z/OS Communications Server 2.5

IP Configuration Guide

SC27-3650-50

Note:

Before using this information and the product it supports, be sure to read the general information under "Notices" on page 1433.

This edition applies to Version 2 Release 5 of z/OS® (5650-ZOS), and to subsequent releases and modifications until otherwise indicated in new editions.

Last updated: 2022-01-25

© Copyright International Business Machines Corporation 2000, 2022.
US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Figures ..... **xxv**

Tables ..... **xxxiii**

About this document ..... **xxxvii**
Who should read this document xxxvii
How this document is organized xxxvii
How to use this document xxxvii
How to contact IBM service xxxvii
Conventions and terminology that are used in this information xxxviii
How to read a syntax diagram xxxix
How to read a syntax diagram xlii
Prerequisite and related information xliv

Summary of changes for IP Configuration Guide ..... **xlix**
Changes made in z/OS Communications Server Version 2 Release 5 xlix
Changes made in z/OS Communications Server Version 2 Release 4 lii
Changes made in z/OS Communications Server Version 2 Release 3 liv

Part 1. Base TCP/IP system **1**

Chapter 1. Overview of z/OS Communications Server 3
TCP/IP protocol stack 4
Multipath channel I/O process 5
Communications Storage Manager 5
Connectivity and gateway functions 5
Network protocol layer 6
Transport layer 7
File systems 7
Application Programming Interfaces 7
TCP/IP socket APIs provided by z/OS Communications Server 8
z/OS UNIX APIs 9

Chapter 2. IP configuration overview 11
IPv6 support 11
z/OS UNIX System Services concepts 11
Overview of data sets and UNIX files 12
Hierarchical file system concepts 12
References to installation data sets 13
Understanding search orders of configuration information 13
Configuration data set naming conventions 13
Configuration files for the TCP/IP stack 23
PROFILE.TCPIP search order 23
TCPIP.DATA search order 24
Configuration files for TCP/IP applications 25
Environment variables 25
MVS-related considerations 28
MVS system symbols 28
Automatic restart manager 29

Logging of system messages....30
Accounting - SMF records....32
Security considerations....34
Nonreusable ASIDs....35
TSO command authorization....36
UNIX System Services security considerations....36
Requirement for an OMVS segment....36
Authorization of TCP/IP started task user ID....36
Other user IDs requiring z/OS UNIX superuser authority....36
BPX.DAEMON FACILITY class profile....37
Program control....38
Defining TCP/IP as a UNIX System Services physical file system....39
Performance considerations....41
Fast path support....42
TCP receive window....43
TCP send window....44
Outbound serialization....45
Considerations for multiple instances of TCP/IP....45
Common INET PFS....45
Port management overview....46
Selecting a stack when running multiple instances of TCP/IP....51
Specifying BPXPRMxx values for a CINET configuration....54
Considerations for Enterprise Extender....54
Considerations for VIPA....55
Considerations for Fast Response Cache Accelerator....56
Considerations for extended address volumes....56
Considerations for networking hardware attachment....57
OSA-Express feature in QDIO mode....57
Steps for converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the IPv4 IPAQENET INTERFACE statement....57
Virtual LAN....60
OSA VLAN....60
OSA routing....61
Relationship of VLAN and primary router....62
Network configuration strategy with VLAN....63
OSA-Express port sharing....67
OSA-Express connection isolation....67
ARP offload and VIPA ARP processing....68
Checksum offload....68
TCP segmentation offload....69
Dynamic LAN idle timer....69
Optimized latency mode....70
QDIO inbound workload queueing....71
Fixed storage considerations for OSA-Express interfaces in QDIO mode....73
Displaying OSA-Express QDIO interface information....74
HiperSockets concepts and connectivity....75
QDIO Accelerator....101
OSA-Express network traffic analyzer trace....104
Synchronization of OSA-Express2 or later diagnostic data....105
Prioritizing outbound OSA-Express data using the Workload Manager service class....105
Fixed storage requirements for OSA-Express QDIO and HiperSockets interfaces....106
Using TEMPIP interfaces....107
Determining the maximum transmission unit....107
Considerations for multiple servers sharing a TCP port....109
Considerations for Common Information Model providers....109
IBM z/OS Container Extensions network overview....110
Required steps before starting TCP/IP....113
Planning your installation and migration....113

Step 1: Install z/OS Communications Server....114
Verifying the initial installation....114
Step 2: Customize z/OS Communications Server....114
Step 3: Configure VMCF and TNF....117
Step 4: Update the VTAM application definitions....120
Step 5: Verify that the required address spaces are active....120
Step 6: Start the TCP/IP address space....121
Step 7: Set up cataloged procedures and configuration data sets....121

Chapter 3. Security....123
Application security....123
TCP/IP resource protection....124
Local user access control to TCP/IP resources using SAF....125
Stack access control....137
Port access control....137
Network access control....140
OSM access control....142
Socket option access control....143
Netstat access control....145
Fast Response Cache Accelerator access control....146
TCP/IP stack initialization access control....146
TCP/IP packet trace service access control....146
TCP connection information service access control....146
zERT information service access control....147
Real-time SMF information service access control....147
TCP/IP OSAENTA trace service access control....147
IPSec network management interface access control....148
CIM provider access control....148
Real-time application-controlled TCP/IP trace NMI access control....148
Syslogd isolation....149
IP filtering....149
Security considerations for the VARY command....150
Multilevel security....150
Configuration data set security....150
Network security principles....151
Cryptography: The foundation of good security....151
End to end security....152
Workload-based security deployment....152
Network security protocols....154
IPSec and VPNs....154
TLS and SSL....157
Application Transparent Transport Layer Security....161
Kerberos....161
Secure Shell (SSH)....162
OSPF authentication....162
SNMPv3....162
Monitoring cryptographic network protection: z/OS encryption readiness technology (zERT)....162
z/OS Encryption Readiness Technology (zERT) Concepts....163
Using z/OS Encryption Readiness Technology....172
Security event reporting: Integrated intrusion detection services....182
Defensive filtering....184
Network security services for the IPSec discipline....185
Network security services for the XMLAppliance discipline....189

Chapter 4. Preparing for IP networking in a multilevel secure environment....191
Understanding multilevel security concepts....191
Multilevel secure networking....191
Nonsecure systems....191

Managed systems....192
Multilevel secure systems....192
z/OS Communications Server TCP/IP stacks on z/OS multilevel secure systems....192
Network security zones....194
IBM zEnterprise System ensemble....194
Where your z/OS systems fit in your network....194
Planning stacks on your z/OS systems....195
Required configuration in a multilevel secure environment....195
Deciding whether to use restricted or unrestricted stacks....196
Steps for configuring global definitions for all stacks....197
Exempting certain users of certain programs from full Network Access Control....198
Configuring stack sysplex features in a multilevel secure environment....198
Defining security labels on other profiles in the SERVAUTH class....199
Planning your multilevel secure network....199
Planning for interactive UNIX System Services users in a multilevel secure environment....200
Steps for creating a separate home directory for each security label....200
Steps for setting stack affinity by security label....201
Host and domain name by security label....201
Planning for applications in a multilevel secure environment....202
Configuring z/OS CS applications in a multilevel secure environment....203
Changing your multilevel secure networking environment....211

Chapter 5. TCP/IP Customization....213
Configuring the syslog daemon....213
Starting and stopping syslogd....219
Configuring syslogd to receive remote messages....225
Offloading log files....228
Setting permissions for log files and directories....229
Configuring syslogd for automatic archiving....230
Using syslogd for z/OS UNIX application programs....231
Usage notes....232
Diagnosing syslogd configuration problems....233
Syslog daemon name/token pair and ECSA storage mapping....234
Configuring TCPIP.DATA....234
Use of TCPIP.DATA and /etc/resolv.conf....234
Creating TCPIP.DATA....235
TCPIP.DATA statements....235
Using MVS system symbols in TCPIP.DATA....236
Configuring PROFILE.TCPIP....237
Changing configuration information....237
Setting up TCP/IP operating characteristics in PROFILE.TCPIP....238
Setting up physical characteristics in PROFILE.TCPIP....250
Setting up reserved port number definitions in PROFILE.TCPIP....259
Setting up the System Authorization Facility server access authorization class (optional)....264
Configuring the local host table (optional)....264
Creating HOSTS.LOCAL site host table....265
Creating /etc/hosts....267
Creating ETC.IPNODES and /etc/ipnodes....267
Verifying your configuration....268
Verifying TCPIP.DATA statement values in the native MVS environment....269
Verifying TCPIP.DATA statement values in the z/OS UNIX environment....269
Verifying PROFILE.TCPIP....269
Verifying interfaces with Ping and Traceroute....269
Verifying local name resolution with TESTSITE....270
Verifying PROFILE.TCPIP and TCPIP.DATA using HOMETEST....270
Verifying your X Window System installation (Optional)....270
Customizing TCP/IP messages....271
Customizing message catalogs....272

Customizing message data sets....276

Chapter 6. Routing....279
Routing terminology....279
General terms....279
Interior Gateway Protocols....280
Route selection algorithm....281
The sample network....282
IPv4 static routing....283
Replaceable static routes....285
IPv6 static routing....285
Replaceable static routes....286
Static routing configuration examples....286
z/OS TCPCS4....286
z/OS TCPCS7....287
IPv4 dynamic routing using OMPROUTE....288
Open Shortest Path First....289
Routing Information Protocol....290
IPv6 dynamic routing using router discovery....291
Multiple routes from router advertisements....291
IPv6 dynamic routing using OMPROUTE....291
IPv6 OSPF protocol....292
IPv6 RIP protocol....292
OMPROUTE configuration....293
Run-time environment....293
Language Environment run-time considerations....294
OMPROUTE tuning considerations....294
Multiple TCP/IP stacks....294
TCP/IP stack routing table management....294
Using RIP, IPv6 RIP, OSPF, and IPv6 OSPF with OMPROUTE....294
Virtual IP addresses....295
Service policy....295
Multiple equal-cost routes....295
Sysplex autonomics....296
Steps for configuring OMPROUTE....296
Starting and controlling OMPROUTE....302
OMPROUTE parameters....302
Controlling OMPROUTE....303
Steps for configuring OSPF and RIP (IPv4 and IPv6)....305
Minimizing the routing responsibility of z/OS Communications Server....324
Preventing futile neighbor state loops during adjacency formation....325
Verification of OMPROUTE IPv4 configuration and state....326
Displaying all OSPF configuration information....326
Displaying information about configured OSPF areas....327
Displaying configuration information about configured OSPF interfaces....327
Displaying information about configured Non-broadcast Multiple Access OSPF interfaces....327
Displaying information about configured OSPF virtual links....327
Displaying information about configured OSPF neighbors....328
Displaying the contents of a single OSPF link state advertisement....328
Displaying statistics and parameters for OSPF areas....329
Displaying the list of AS external advertisements....329
Displaying a list of non-AS external advertisements....329
Displaying current, run-time statistics and parameters for OSPF interfaces....329
Displaying current, run-time statistics and parameters for a specific OSPF interface....330
Displaying current, run-time statistics and parameters for OSPF neighbors....330
Displaying current run-time statistics and parameters for a specific OSPF neighbor....330
Displaying routes to other routers that have been calculated by OSPF....331
Displaying the number of LSAs currently in the link state database....331

Displaying statistics generated by the OSPF routing protocol....331
Displaying all of the RIP configuration information....331
Displaying information about configured RIP interfaces....332
Displaying the routes to be unconditionally accepted....332
Displaying current run-time information about RIP interfaces....332
Displaying current run-time information about a specific RIP interface....332
Displaying the global RIP filters....333
Displaying the routes in the OMPROUTE main routing table....333
Displaying the routes to a specific destination in the main routing table....334
Displaying the routes in all OMPROUTE IPv4 policy-based routing tables....334
Displaying the routes in an OMPROUTE IPv4 policy-based routing table....335
Displaying the routes to a specific destination in an IPv4 policy-based routing table....336
Displaying all of the generic configuration information....336
Displaying information about configured generic interfaces....336
Displaying current run-time information about generic interfaces....336
Verification of OMPROUTE IPv6 configuration and state....337
Displaying all IPv6 OSPF information....337
Displaying IPv6 OSPF area statistics and parameters....337
Displaying IPv6 OSPF interface statistics and parameters....338
Displaying statistics and parameters for a specific IPv6 OSPF interface....338
Displaying IPv6 OSPF virtual link statistics and parameters....338
Displaying statistics and parameters for a specific IPv6 OSPF virtual link....338
Displaying IPv6 OSPF neighbor statistics and parameters....339
Displaying statistics and parameters for a specific IPv6 OSPF neighbor....339
Displaying IPv6 OSPF link state database statistics....339
Displaying IPv6 OSPF link state advertisement....340
Displaying IPv6 OSPF external advertisements....340
Displaying IPv6 OSPF area link state database....340
Displaying IPv6 OSPF router routes....341
Displaying IPv6 OSPF routing protocol statistics....341
Displaying all of the IPv6 RIP information....342
Displaying information about IPv6 RIP interfaces....342
Displaying information about a specific IPv6 RIP interface....342
Displaying the routes to be unconditionally accepted by IPv6 RIP....342
Displaying the global IPv6 RIP filters....343
Displaying the routes in the OMPROUTE IPv6 main routing table....343
Displaying the routes to a specific destination in the IPv6 main routing table....344
Displaying the routes in all OMPROUTE IPv6 policy-based routing tables....344
Displaying the routes in an OMPROUTE IPv6 policy-based routing table....345
Displaying the routes to a specific destination in an IPv6 policy-based routing table....346
Displaying all of the IPv6 generic information....346
Displaying information about IPv6 generic interfaces....346
Displaying information about a specific IPv6 generic interface....347
Sample OMPROUTE configuration files....347
Policy-based routing....349
Options for configuring policy-based routing....350
Routing policy configuration....351
Getting started with policy-based routing....354
Configuring policy-based routing....354
Considerations for using policy-based routing with IP security....356
Considerations for mixed routing environments....357
Use of static routing with OMPROUTE....357
Use of IPv6 static routing with router advertisements....357
Use of policy-based routing with static or dynamic routing....358
Verifying static, dynamic, and policy-based routing....358
Verifying connections with Netstat, Ping, and Traceroute....358

Chapter 7. Virtual IP Addressing....361

Terminology....361
Introduction to VIPA....361
Moving a VIPA (for TCP/IP outage)....363
Static VIPAs, dynamic VIPAs, distributed DVIPAs....363
Using static VIPAs....364
Steps for configuring static VIPAs for a z/OS TCP/IP stack....365
Steps for converting from IPv4 VIRTUAL DEVICE, LINK, and HOME definitions to the IPv4 VIRTUAL INTERFACE statement....367
Configuring static VIPAs for Enterprise Extender....368
Considerations when using static VIPAs with IPv6....368
Planning for static VIPA takeover and takeback....368
Using dynamic VIPAs....369
Configuring DVIPA support....369
Planning for dynamic VIPA takeover....369
Different application uses of IP addresses and DVIPAs....372
Configuring dynamic VIPAs....372
Configuring the multiple application-instance scenario....373
Configuring the unique application-instance scenario....373
Configuring application-instance DVIPAs for IBM z/OS Container Extensions (zCX)....380
Choosing which form of dynamic VIPA support to use....381
Configuring distributed DVIPAs - sysplex distributor....382
Manually quiescing DVIPA sysplex distributor server applications....385
Route selection for distributing packets....385
Generic routing encapsulation....387
Dynamic port assignment....388
Sysplex-wide source VIPA....388
GLOBALCONFIG EXPLICITBINDPORTRANGE....391
Timed affinities....393
Sysplex-Wide Security Associations....396
Resolution of dynamic VIPA conflicts....399
Restart of the original VIPADEFINE TCP/IP after an outage....399
Movement of unique application-instance (BIND)....400
Movement of a unique APF-authorized application instance (ioctl)....404
Same dynamic VIPA for VIPADEFINE and BIND(), SIOCSVIPA or SIOCSVIPA6 ioctl, or MODDVIPA utility....404
Dynamic VIPA creation results....405
TIER1, TIER2, and CPCSCOPE keyword DVIPA contention resolution....409
IPv6 considerations....414
VIPARANGE....414
VIPADEFINE and VIPABACKUP....414
Unique application-instance scenario and IPv6-enabled applications....414
VIPAs, OSA-Express QDIO, and Spanning Tree Protocol....415
Mixture of types of dynamic VIPAs within subnets....415
MVS failure and sysplex failure management....416
Applications and dynamic VIPAs....416
Configuring VIPAs for activation with VIPABACKUP....417
Example of configuring dynamic and distributed VIPAs....418
Verifying the DVIPAs in a sysplex....420
Using Netstat support to verify dynamic VIPA configuration....423
Verifying sysplex distributor workload....427
Dynamic VIPAs and routing protocols....431
IPv4 considerations for OMPROUTE....431
IPv4 considerations for Routing Information Protocol....432
IPv6 considerations....432

Chapter 8. TCP/IP in a sysplex....435
Connectivity in a sysplex....436
Sysplex subplexing....436

Dynamic XCF....440
Network interfaces monitoring....451
Sysplex problem detection and recovery....452
Target server connection setup responsiveness monitoring....468
Workload balancing....468
Single systemwide image....469
Horizontal growth....469
Ease of management....469
Internal load balancing solutions....470
Sysplex-aware external load balancing solutions....470
External IP workload balancing solutions....470
Choosing a load balancing solution....471
Sysplex distributor....475
BASEWLM - Distribution using WLM system weights....475
SERVERWLM - Distribution using WLM server-specific weights....476
Choosing between the BASEWLM and SERVERWLM distribution methods....479
BASEWLM and SERVERWLM display example....479
WEIGHTEDACTIVE - Distribution based on active connection load....480
Choosing between RoundRobin and WeightedActive distribution....481
Hot standby distribution....481
Timed affinity....484
SHAREPORT....484
QDIO Accelerator....485
QDIO inbound workload queueing....485
Optimizing local connections....485
Policy interactions....485
Sysplex distribution optimizations for multi-tier z/OS workloads....487
Sysplex distributor optimization with the OPTLOCAL keyword....487
Sysplex distributor enhanced workload distribution for z/OS multi-tier, OPTLOCAL configurations....489
Sysplex distributor enhanced workload distribution for z/OS multi-tier, OPTLOCAL configurations with CPC affinity....490

Chapter 9. TCP/IP in an ensemble....493
Steps for configuring an interface for the intraensemble data network (CHPID type OSX)....494
HiperSockets connectivity to the intraensemble data network....494
Operating and managing IEDN-enabled HiperSockets interfaces....496
Performance considerations for the IEDN-enabled HiperSockets function....497
Steps for enabling HiperSockets access to the intraensemble data network....498
Steps for enabling IPv6 on a stack for access to the intranode management network....499
Steps for using the intranode management network (CHPID type OSM)....499
Routing considerations for the intraensemble data network....500
OMPROUTE considerations for the intraensemble data network....501
Sysplex distributor considerations for the intraensemble data network....501
Multilevel security and network access control considerations....501

Chapter 10. Shared Memory Communications....503
Shared Memory Communications over Remote Direct Memory Access....503
Remote Direct Memory Access over Converged Ethernet....504
Comparing 10 GbE RoCE Express feature environments....504
RoCE Express2 feature environment....507
Shared Memory Communications - Direct Memory Access....507
Shared Memory Communications concepts....507
Rendezvous processing....507
Shared Memory Communications links and link groups....510
SMC memory buffers....514
SMC-R staging buffers....515
SMC filters....515

Shared Memory Communications terms....515
Using Shared Memory Communications....518
Configuration considerations for Shared Memory Communications....518
Setting up the environment for Shared Memory Communications over RDMA....529
Configuring Shared Memory Communications over RDMA....530
Setting up the environment for Shared Memory Communications - Direct Memory Access....532
Configuring Shared Memory Communications - Direct Memory Access....532
Shared Memory Communications multiple IP subnet support (SMCv2)....534
SMC Enterprise ID....534
EID Format....537
User Defined EID (UEID)....537
Multiple EIDs....537
System EID (SEID)....538
Using UEID or SEID....538
SMC filters....539
Using SMC filters....539
Dynamic selection of SMC version and SMC type....540
SMC-Dv2....540
ISMv2....540
Enabling SMC-Dv2....540
SMC-Dv2 and ISM VCHID Association using PNetIDs....541
SMC-Rv2....547
RoCEv2....547
Enabling SMC-Rv2....548
OSA and RoCEv2 device association and physical connectivity....549
SMC-Rv1 migration considerations....552
Sharing or reusing resources: SMCRIPADDR and PFID....552
SMC-Rv2 resilience, high availability with RoCEv2 and IP routing....552
SMC-Rv2 link groups....552
SMC interactions with other z/OS Communications Server functions....555
Sysplex distributor....555
Security functions....556
Intrusion detection services (IDS)....556
TCP keepalive....557
TCP application data transfer options....558
Packet trace....558
SMC-R RoCE maximum transmission unit....558
Managing SMC communications....558
Managing your "RoCE Express" interfaces....559
Managing your ISM interfaces....560
Displaying SMC information....560
Monitoring SMC information....562
VTAM displays and tuning statistics....563
Stopping SMC-R....564
Stopping SMC-D....565

Part 2. Server applications....567

Chapter 11. Accessing remote hosts using Telnet....569
The TN3270E Telnet server....569
Steps for starting the TN3270E Telnet server....569
Managing Telnet....577
Telnet diagnostic tools....580
Telnet configuration data set customization details....585
Configuring the z/OS UNIX Telnet server....655
Installation information....655
Environment variables....656

Starting, stopping, and administration of z/OS UNIX Telnet....656
otelnetd....659
SMF record handling....662
BPX.DAEMON considerations....662
Kerberos....662

Chapter 12. Transferring files using FTP....663
Configuring PROFILE.TCPIP for FTP....664
Configuring ETC.SERVICES....665
Configuring /etc/syslog.conf....665
Configuring the FTPD cataloged procedure....665
Security for the FTP server....666
Defining environment variables for the FTP server (optional)....673
Configuring FTP with multiple TCP/IP stacks....675
Configuring TCPIP.DATA for FTP....675
Configuring FTP.DATA....676
Optionally configuring user-level server options using FTPS.RC....676
Data set attributes....676
Specifying attributes for new MVS data sets....677
Translation of data....680
z/OS UNIX named pipes....680
FTP code page conversion....680
Master catalog access....683
Customizing FTP message catalogs....683
Steps for creating a message catalog from the shipped catalog and preserving its timestamp....683
Accounting....685
Configure the FTP server for SMF (optional)....685
Customizing Transport Layer Security and Kerberos security....686
Steps for customizing the FTP server for TLS....686
Steps for customizing the FTP server for Kerberos....691
Steps for customizing the FTP client for TLS....695
Steps for customizing the FTP client for Kerberos....700
Port 990....702
Steps for migrating the FTP client to use AT-TLS....703
Traversing firewalls with SSL/TLS secure FTP....705
Db2 and JES....708
Configuring the optional FTP user exits....708
The FTPSMFEX user exit (for the FTP server)....708
The FTCHKIP user exit (for the FTP server)....709
The FTCHKPWD user exit (for the FTP server)....709
The FTCHKCMD user exit (for the FTP server)....709
The FTCHKJES user exit (for the FTP server)....710
The FTPOSTPR user exit (for the FTP server)....710
The EZAFCCMD user exit (for the FTP client)....711
The EZAFCREP user exit (for the FTP client)....712
Customizing the FTP-to-JES interface for JESINTERFACELevel 2 (optional)....713
Configuring the FTP server for anonymous FTP (optional)....714
Creating an anonymous directory structure in the z/OS UNIX file system....715
Configure the welcome banner page, login, and directory message (optional)....718
Using magic cookies to represent information....718
Configuring the FTP server to log session (user ID) activity....719
Configuring to send detailed login failure replies to an FTP client (optional)....720
Install the SQL query function (optional) and access the Db2 modules....720
Accessing Db2 modules....721
FTP.DATA updates for SQL query function....722
Verifying the FTP server....722
Verifying the FTP client....723

Verifying FTP.DATA statements ... 724
Verifying anonymous, banner, and other optional configuration information ... 726
Verifying the FTP-JES interface (optional) ... 726

Chapter 13. The resolver ... 729
DNS overview ... 729
Domain names ... 730
Domain name servers ... 730
Resolvers ... 733
Querying name servers ... 735
Recommended reading ... 736
Resolver API calls ... 737
Starting the resolver ... 737
The default resolver settings ... 738
Customizing the resolver ... 738
The resolver setup file ... 738
The resolver address space ... 745
Managing the resolver address space ... 747
Steps for manually restarting the resolver ... 748
Steps for applying an interim fix to the resolver ... 748
IPv6 name servers and the resolver ... 749
Resolver functions ... 750
Resolver caching ... 750
Monitoring the responsiveness of Domain Name System name servers ... 761
Extension Mechanisms for DNS standards and the resolver ... 770
Resolver configuration files ... 771
z/OS XL C/C++ environment variables for configuration files ... 775
Search orders used in the z/OS UNIX environment ... 777
Search orders used in the native MVS environment ... 781

Chapter 14. Policy-based networking ... 787
Policy types and infrastructure overview ... 787
Configuration files and policy definition files ... 789
Managing changes to configuration files and policy definition files ... 789
Storing configuration files and policy definition files ... 790
Steps for managing policy changes ... 791
Policy infrastructure components ... 793
TCP/IP stack ... 793
Policy Agent ... 793
Traffic regulation management daemon ... 797
IKE daemon ... 797
Network security services daemon ... 797
Defense Manager daemon ... 797
SNMP Network SLAPM2 subagent ... 798
Sample policy infrastructure ... 798
Policy sample files ... 800
Policy types ... 801
QoS policy ... 802
IDS policy ... 802
IPSec policy ... 802
AT-TLS policy ... 803
Policy-based routing policy ... 804
zERT policy-based enforcement (ZERT) policy ... 805
Policy configuration files ... 806
Steps for configuring the Policy Agent ... 808
Step 1: Configure general information ... 808
Step 2: Configure Policy Agent as a policy server ... 811
Step 3: Configure Policy Agent as a policy client ... 816

Step 4: Configure policies in Policy Agent configuration files ... 817
Step 5: Configure Policy Agent to use the LDAP server using the ReadFromDirectory statement ... 817
Step 6: Configure Policy Agent for import services ... 818
Step 7: Configuring Policy Agent to automatically monitor applications ... 820
Add TLS/SSL to Policy Agent connections ... 822
Starting and stopping the Policy Agent ... 823
AUTOLOG considerations ... 823
Specifying environment variables ... 824
Main configuration file search order ... 825
Other considerations when starting the Policy Agent ... 825
Stopping the Policy Agent ... 827
Refreshing policies ... 827
FLUSH and PURGE considerations ... 827
Switching between local and remote policies ... 829
Verifying that policies are correctly defined and functioning properly ... 830

Chapter 15. Quality of service ... 831
Differentiated Services policies ... 831
Integrated Services policies ... 833
Sysplex distributor policies ... 833
QoS-specific Policy Agent functions ... 833
Sysplex distributor policy performance monitoring configuration ... 834
Policy performance collection configuration ... 835
IPv4 type of service or IPv6 traffic class mapping configuration ... 836
Options for configuring QoS ... 837
Option 1: Use the IBM Configuration Assistant for z/OS Communications Server ... 837
Option 2: Manual configuration ... 838
Specifying the QoS configuration file based on Policy Agent role ... 838
Defining policies in a Policy Agent configuration file ... 838
Differentiated Services policy examples ... 839
RSVP policy example ... 840
Sysplex distributor policy example ... 840
Defining policies using LDAP ... 841
RSVP ... 841
Configuring the RSVP agent ... 842
Starting and stopping RSVP ... 842
SNMP Network SLAPM2 (nslapm2) performance monitor ... 843
Configuring the Network SLAPM2 subagent ... 843
Starting and stopping the Network SLAPM2 subagent ... 844
Verification ... 845
Verifying that the policies are installed in the TCP/IP stacks ... 845
Verifying that the expected traffic is mapping to the correct QoS policies ... 845
Verifying that the sysplex distributor policy functions are working correctly ... 846
Monitoring performance and tuning policies ... 846
Using pasearch ... 846
Using the Network SLAPM2 MIB to monitor policies ... 846

Chapter 16. Intrusion detection services ... 851
Scan policies ... 851
ICMP scans ... 853
ICMPv6 scans ... 853
UDP port scans ... 854
TCP port scans ... 854
Attack policies ... 855
Traffic regulation policies ... 861
Traffic regulation policies for TCP ports ... 861
Traffic regulation policies for UDP ports ... 862

Options for configuring IDS....863
Option 1: Use the IBM Configuration Assistant for z/OS Communications Server....863
Option 2: Manual configuration....864
Specifying the IDS configuration file based on Policy Agent role....864
Defining IDS policies....864
IDS policy definition considerations....864
IDS scan policy example....867
IDS attack policy examples....869
Traffic Regulation policy examples....876
Verification....878
Are the correct policies active?....879
Is the expected traffic mapping to the correct policies?....879
Are the IDS policy functions working correctly?....879
TRMD....879
Running TRMD as a started task....880
Running TRMD from the z/OS UNIX shell....881
Stopping TRMD....881
trmdstat....881
Defensive filtering....882

Chapter 17. IP security....883
Terms and concepts for IP security....883
Terminology conventions for IP security....887
Commands used to administer IP security....887
Overview of using IP security....888
FIPS 140 mode and IP security....889
Configuring IP security....892
IP filtering....893
Filter rules and actions....893
Filtering criteria in an IP packet....894
Additional filtering criteria based on protocol....894
Additional filtering criteria based on network attributes....895
IP traffic patterns....897
Routed traffic and fragmented packets....897
Conditionally controlling IP filters....898
Special considerations when using IP security for IPv6....898
Neighbor discovery and multicast listener discovery....898
Stateless address autoconfiguration....899
IPv6-specific protocols....899
IPv6 address types....899
IPv6 extension headers....899
Considerations for IPv6 OSPF security....899
Default IP filter policy and IP security policy....901
Modifying the default IP filter policy....902
IP filter logging....919
IP filter discard action....919
Data encryption and authentication - IPSec....919
AH and ESP protocols....920
IPSec and symmetric key management....925
Manual key management....925
Dynamic key management - IKE and IPSec negotiations....926
IPSec and network address translation devices....932
Dynamic structures used to map Security Associations....933
Steps for preparing the z/OS system for IP security....935
IP security policy configuration....939
Overview of configuring IP security policy....939
Steps for configuring local IP security policy using only a common IP security configuration file....942

Steps for configuring remote IP security policy using only a common IP security configuration file....942
Steps for configuring local IP security policy using only a stack-specific IP security configuration file....943
Steps for configuring remote IP security policy using only a stack-specific IP security configuration file....944
Steps for configuring local IP security policy using both a stack-specific file and a common file....945
Steps for configuring remote IP security policy using both a stack-specific file and a common file....945
Component policies of IP security policy configuration files....946
Quick start using IP filtering and IPSec host-to-host....957
Steps for configuring IP security policy....972
Configuring specific security models....974
Configuring the IKE daemon....1041
Multiple TCP/IP stacks....1041
Run-time environment....1041
Language Environment run-time considerations....1042
IKE daemon configuration source information....1042
Policy Agent considerations....1042
Using network security services....1042
Certificate revocation checking....1044
Steps for configuring the IKE daemon....1045
Starting the IKE daemon....1048
Stopping the IKE daemon....1049
Controlling the IKE daemon....1050
Verifying policy installation....1050
Console messages....1050
Displaying TCP/IP configuration....1050
Displaying active filters with the ipsec command....1051
Displaying Security Associations with the ipsec command....1067
Displaying filter rules with the pasearch command....1069
Verifying filter action....1069
Security Associations....1073
Activating a Security Association....1073
Verifying the activation of a Security Association....1074
Verifying the use of an active Security Association....1074
Refreshing Security Associations....1074
Deactivating Security Associations....1075
Modifying active IP security policy....1076
IP security policy files....1076
Policy Agent image configuration files....1076
Policy Agent main configuration file....1077
Active Security Associations and the ipsec -f default command....1077
Displaying NSS client information....1078
Sysplex-Wide Security Associations and IP security....1079
NAT traversal and Sysplex-Wide Security Associations....1080
FIPS 140 mode and Sysplex-Wide Security Associations....1080
Shadow Security Associations....1081
Sample IP security policy files....1082

Chapter 18. Network security services....1083
Terms and concepts for network security services....1083
Network security services overview....1084
NSS IPSec discipline overview....1084
NSS XMLAppliance discipline....1085
Preparing to provide network security services....1085
Steps for authorizing resources for NSS....1086

NSS server certificate label naming considerations....1092
NSS client authorization example....1093
NSS server configuration considerations....1094
Using hash and URL certificate encoding types....1099
Creating certificate bundles....1100
Controlling the NSS server....1102
NSS server failover considerations....1103
NSS server capacity considerations....1104
NSS server certificate revocation support....1104
Managing network security services....1104

Chapter 19. Defensive filtering....1107
Global and stack-specific defensive filters....1108
Defensive filter names....1109
Defensive filter modes....1109
Allowing administrative access....1110
Filter-match logging....1110
TRMD....1111
Disabling defensive filters for a single stack....1111
Relationship between intrusion detection services and defensive filters....1111
Comparison of IP security filters and defensive filters....1112
The DMD run-time environment....1114
The DMD and Language Environment run-time options....1115
Enabling defensive filtering....1115
Enabling the IP security function....1115
Steps for configuring the DMD....1116
Steps for authorizing resources for the DMD and the ipsec command....1118
Starting the DMD....1118
Stopping the DMD....1119
Using the DMD MODIFY command....1119

Chapter 20. Application Transparent Transport Layer Security data protection....1121
AT-TLS configuration in PROFILE.TCPIP....1121
TCP/IP stack initialization access control....1122
Options for configuring AT-TLS security....1122
Option 1: Use the IBM Configuration Assistant for z/OS Communications Server....1122
Option 2: Manual configuration....1123
Specifying the AT-TLS configuration file based on Policy Agent role....1123
AT-TLS policy configuration....1123
AT-TLS rules....1124
AT-TLS actions....1124
Getting started with AT-TLS....1126
Configuring the server system....1126
Configuring the client systems....1128
Steps for starting AT-TLS and verifying its operation....1130
Application compatibility with AT-TLS....1131
Policy considerations....1131
Reusable objects....1131
Common AT-TLS configuration file....1131
Exempting specific connections from AT-TLS....1131
Action refresh....1132
Achieving the basic level of security....1132
Picking the handshake roles....1132
Specifying the key ring....1133
Configuring more sophisticated security....1134
Cipher suite specification....1134
Protocol versions....1134
Using TLSv1.3 protocol support....1135

Certificate validation..... 1135
FIPS 140-2 support.....1136
Certificate revocation support..... 1136
Encryption key refresh..... 1136
Additional security customization considerations.....1137
Handshake timer..... 1137
Diagnostic traces..... 1137
Diagnosis considerations..... 1138
TLS function negotiation..... 1138
Session caching..... 1139
AT-TLS access control considerations.....1140
Application model considerations.....1140
Client application model..... 1140
Server application model..... 1141
Forked server application model..... 1142
CICS transaction model..... 1143
Advanced application considerations..... 1144
AT-TLS aware application considerations.....1144
AT-TLS controlling application considerations..... 1144
Secondary connection application model..... 1145

Chapter 21. z/OS Load Balancing Advisor..... 1147
Steps for preparing to use the z/OS Load Balancing Advisor..... 1148
Step 1: Consider whether to use TLS/SSL (using AT-TLS on z/OS)..... 1149
Step 2: Evaluate TCP/IP workloads to be load balanced and select a load balancing solution (optional)..... 1150
Step 3: Decide who will have authority to start the Advisor (optional)..... 1150
Step 4: Decide who will have authority to start the Agents (optional)..... 1151
Step 5: Authorize the Agents to use WLM services..... 1151
Step 6: Determine how the Advisor and agent are to interact in a subplexing environment (optional)..... 1152
Steps for configuring the z/OS Load Balancing Advisor.....1152
Step 1: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)..... 1153
Step 2: Configure and start syslogd..... 1155
Step 3: Configure one Advisor per sysplex..... 1156
Step 4: Configure one Agent per z/OS system in the sysplex..... 1161
Step 5: Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run (optional)..... 1163
Step 6: Customize WLM policies for the Advisor and Agents (optional).....1168
Step 7: Configure the external load balancers..... 1168
Steps for starting the z/OS Load Balancing Advisor..... 1171
Step 1: Start the TCP/IP stacks that the Advisor and the Agents will use.....1171
Step 2: Start the target applications that will be the targets of load balancing..... 1171
Step 3: Start one Agent on each sysplex system you want to participate in this method of workload balancing..... 1171
Step 4: Start the one instance of the Advisor in the sysplex.....1172
Step 5: Start the load balancers..... 1172
Verifying that the Advisor system is functioning correctly (optional).....1172
Operating the z/OS Load Balancing Advisor..... 1173
Changing the logging level of the Advisor and Agents..... 1173
Interpreting Agent and Advisor display information.....1173
Stopping or resuming workload distribution to particular members (QUIESCE and ENABLE)1186
z/OS Load Balancing Advisor configuration example..... 1188
Load balancer configuration details.....1188
Advisor configuration details..... 1189
Agent configuration file on SYSB.....1190
Agent configuration file on SYSA.....1191

Customization of PROFILE.TCPIP ... 1191
Example displays ... 1193

Chapter 22. Automated domain name registration ... 1195
System overview ... 1195
Interaction with name servers ... 1196
Interaction with the z/OS Load Balancing Advisor ... 1196
Enabling TLS/SSL for ADNR ... 1197
Steps for configuring automated domain name registration ... 1197
Step 1: Decide which sysplex resources should be managed by ADNR ... 1198
Step 2: Decide on one or more domain names to be managed by ADNR ... 1198
Step 3: Decide which name server or name servers are to be managed by ADNR ... 1199
Step 4: Configure the selected name servers to be the primary master name servers for the domain names that ADNR is to manage ... 1199
Step 5: Delegate the domain names to be managed by ADNR to the selected name servers from the parent domain's name server ... 1199
Step 6: Configure the z/OS Load Balancing Advisor function ... 1200
Step 7: Define security server profiles for ADNR ... 1200
Step 8: Configure ADNR to automatically restart in case of application or system failure (optional) ... 1201
Step 9: Configure and start syslogd (optional, but required to have ADNR write log messages and trace data to syslogd) ... 1202
Step 10: Configure one ADNR application per sysplex ... 1202
Step 11: Customize the TCP/IP profiles of the TCP/IP stacks on which ADNR and the LBA applications are to run (optional) ... 1206
Step 12: Start the TCP/IP stacks on which ADNR and the LBA applications are to run ... 1206
Step 13: Start the z/OS Load Balancing Advisor and Agent ... 1207
Step 14: Start the target applications that are to be managed by ADNR ... 1207
Step 15: Start the ADNR application ... 1207
Step 16: Verify that the ADNR system is functioning correctly (optional) ... 1207
z/OS Load Balancing Advisor configuration considerations ... 1208
Connectivity considerations ... 1208
Near real-time availability information of sysplex resources ... 1209
z/OS Load Balancing Advisor and Agent operational considerations ... 1209
Advisor operational considerations ... 1209
Agent operational considerations ... 1209
Name server configuration considerations ... 1210
Initial zone configuration ... 1210
Authorizing dynamic updates ... 1210
Updates to an ADNR-managed zone ... 1210
Authorizing zone transfers ... 1211
Limiting the duration of an outbound zone transfer ... 1212
Limiting the total number of simultaneous outbound zone transfers ... 1212
The .digrc file ... 1212
Split DNS (views) ... 1212
Zone transfer formats ... 1213
ADNR configuration considerations ... 1214
Changing the ADNR configuration file ... 1214
Maintaining zone data integrity ... 1214
Steps for using the ADNR application in a sysplex subplexing environment ... 1215
Step 1: Plan how the new subdomains representing each subplex will fit into your DNS hierarchy ... 1216
Step 2: Configure the name servers that will be updated for the new subplex domains ... 1217
Step 3: Define and configure one Advisor per subplex ... 1217
Step 4: Update the Agent configuration files to communicate with the Advisor running in its subplex ... 1218
Step 5: Define one ADNR application per subplex ... 1218

Step 6: Assign the host_group and server_group statements from the sysplex ADNR configuration to the correct subplex domains....1218
Step 7: Configure the new ADNR instances to update the name server and zone for its subplex....1219
Step 8: Configure the new ADNR instances to communicate with the subplex Advisor....1219
Step 9: Update resolver configuration files (optional)....1219
Step 10: Start the TCP/IP stacks, Advisor, Agent, ADNR, and target applications that are to be managed by ADNR....1219
Step 11: Verify that each subplex ADNR is functioning correctly....1219
Operating ADNR....1219
Changing the logging level of ADNR....1220
Changing the ADNR configuration dynamically....1220
Interpreting ADNR display information....1220
Diagnosing problems....1220
ADNR configuration example....1221
ADNR display examples....1228

Chapter 23. Simple Network Management Protocol....1237
SNMP overview....1237
Network management application....1238
SNMP protocols....1238
SNMP agent....1239
SNMP subagents....1240
Key generation commands....1242
Distributed Protocol Interface....1242
Trap forwarder daemon....1242
Processing an SNMP request....1242
Deciding on SNMP security needs....1243
Community-based security....1243
User-based security....1243
Decide on your security needs - community-based or user-based....1244
Step 1: Configure the SNMP agent....1245
Provide TCP/IP profile statements....1245
Provide community-based security and notification destination information....1246
Provide community-based and user-based security and notification destination information1249
Migrating community-based configuration to SNMPD.CONF format....1251
Provide secure access to agent from subagents....1252
Allowing subagents with duplicate identifiers to connect....1253
Provide MIB object configuration information....1253
Common INET considerations....1254
Start the SNMP agent....1255
Sample JCL procedure for starting OSNMPD from MVS....1255
Starting OSNMPD from z/OS UNIX....1255
Step 2: Configure the SNMP commands....1255
Configure the z/OS UNIX snmp command....1256
Configure the NetView SNMP command....1258
Step 3: Configure the SNMP subagents....1261
TCP/IP subagent configuration....1261
Step 4: Configure the Open Systems Adapter support....1261
OSA/SF prerequisites....1263
Required TCP/IP profile statements....1263
Subagent connection to OSA/SF when there are multiple TCP/IP instances....1264
Step 5: Configure the trap forwarder daemon....1265
Provide PROFILE.TCPIP statements....1265
Provide trap forwarder configuration information....1265
Starting and stopping the trap forwarder daemon....1266

Chapter 24. Remote print server....1269

Configuring the Remote Print Server....1269
Step 1: Configuring PROFILE.TCPIP for LPD....1269
Step 2: Updating the LPD server cataloged procedure....1270
Step 3: Updating the LPD server configuration data set....1271
Step 4: Creating a banner page (optional)....1271

Chapter 25. Remote procedure calls....1273
Steps for configuring the PORTMAP address space....1273
Step 1: Configuring PROFILE.TCPIP for PORTMAP....1273
Step 2: Configuring security server (or RACF equivalent) items....1274
Step 3: Updating the PORTMAP cataloged procedure....1274
Step 4: Defining the data set for well-known procedure names....1274
Starting the PORTMAP address space....1275
Steps for configuring the z/OS UNIX PORTMAP address space....1275
Step 1: Configuring PROFILE.TCPIP for UNIX PORTMAP....1275
Step 2: Updating the PORTMAP cataloged procedure....1276
Starting the PORTMAP address space....1276
Steps for configuring the rpcbind address space....1276
Step 1: Configuring the PROFILE.TCPIP data set for rpcbind....1277
Step 2: Configuring security server (or RACF equivalent) items....1277
Step 3: Updating the RPCBIND cataloged procedure....1278
Step 4: Updating the /etc/services file....1278
Step 5: Configure SYS1.PARMLIB for rpcbind....1278
Starting the rpcbind address space....1279
Steps for configuring the NCS interface....1280
Step 1: Configuring PROFILE.TCPIP for NCS....1281
Step 2: Updating the NRGLBD cataloged procedure....1281
Step 3: Updating the LLBD cataloged procedure....1281

Chapter 26. Mail on z/OS....1283
Configuring the CSSMTP application....1283
Terms and concepts....1284
Setting up CSSMTP....1286
Customizing the CSSMTP configuration file to try mail again....1289
Customizing the CSSMTP configuration file to handle undeliverable mail....1290
Steps for granting authority to start CSSMTP....1291
Security for CSSMTP....1292
Steps for using Transport Layer Security for CSSMTP....1293
Steps for configuring SMF records for CSSMTP (optional)....1295
Monitoring CSSMTP....1297
sendmail to CSSMTP bridge....1297
Setting up the sendmail bridge....1298
Steps for creating and customizing the configuration file....1298
Security considerations for the sendmail bridge....1299
Configuring popper....1299
Update the /etc/services file....1300
Update the /etc/inetd.conf file....1300
Create the directory for the temporary maildrop file....1300
Start inetd....1300
Correct connection....1300
Popper command - administering received mail....1301

Chapter 27. TIMED daemon....1303
Starting TIMED from the z/OS shell....1303
Starting TIMED as a procedure....1303

Chapter 28. SNTP daemon....1305
Steps for starting SNTPD from the z/OS UNIX shell....1305

Steps for starting SNTPD as a procedure....1306
Stack affinity....1307

Chapter 29. Remote Execution....1309
UNIX REXEC....1309
TSO REXEC....1309
Configuring the TSO Remote Execution server....1309
Step 1: Configuring PROFILE.TCPIP for TSO Remote Execution server....1310
Step 2: Determine whether Remote Execution client will send REXEC or RSH commands....1310
Step 3: Permit remote users to access MVS resources (optional)....1310
Step 4: Update the TSO Remote Execution cataloged procedure....1311
Step 5: Create a user exit routine (optional)....1311
Step 6: Permit access to JESSPOOL files....1313
Configuring the z/OS UNIX Remote Execution servers....1313
Files for z/OS UNIX REXECD....1313
Files for z/OS UNIX RSHD....1314
Considerations for Kerberos authentication....1314
Setting up the z/OS UNIX RSHD installation exit....1314
Configuring TSO and z/OS UNIX Remote Execution servers to use the same port....1315

Chapter 30. Express logon services with the Digital Certificate Access Server....1317
Express Logon Feature....1317
Web Express Logon....1317
Using the DCAS server interface for your logon solutions....1317
What DCAS provides....1317
Customizing DCAS for TLS/SSL....1318
Transport Layer Security (TLS) terms....1320

Chapter 31. Miscellaneous server....1323
Discard protocol....1323
Echo protocol....1323
Character generator protocol....1323
Configuring the MISC server....1324
Step 1: Configuring PROFILE.TCPIP for the MISC server....1324
Step 2: Updating the MISC server cataloged procedure....1324

Appendix A. Setting up the InetD configuration file....1327

Appendix B. TLS/SSL security....1329
Secure Socket Layer overview....1329
Server authentication....1329
Client authentication....1330
Encryption algorithms....1332
Enable CSFSERV resources....1332

Appendix C. Express Logon Feature....1335
Configuring RACF services for Express Logon....1336
An example of configuring the Express Logon components....1337
Configuring the Host On Demand Telnet client....1337
Configuring the z/OS TN3270E Telnet server (two-tier solution)....1337
Configuring the middle-tier Telnet server (IBM Communications Server for Windows example) 1337

Appendix D. Using HCD....1339

Appendix E. Steps for preparing to run IP security....1361
Step 1: Setting appropriate UNIX System Services parameters....1361
Step 2: Ensure the IKE daemon's user ID has enough system resources....1361

Virtual storage capacity 1361
System message queue capacity 1362
UDP queue capacity 1362
Step 3: Authorizing the IKE daemon to the external security manager 1362
Steps for authorizing the IKE daemon to RACF 1362
Step 4: Authorizing the ipsec command to the external security manager 1363
Steps for authorizing the ipsec command to RACF 1363
Step 5: Authorizing IP security to ICSF (optional) 1364
Steps for setting up profiles in the CSFSERV resource class 1365
Step 6: Setting up the IKE daemon for digital signature authentication (optional) 1367
Steps for setting up the IKE daemon for digital signature authentication when the native certificate service is used 1370
Steps for setting up the IKE daemon for digital signature authentication using the certificate service of an NSS server 1372
IPSec certificate management 1373

Appendix F. Using an LDAP server for policy definitions 1377
Policy object model overview 1377
Overview of the object classes 1381
Considerations for defining LDAP objects 1386
Policy Agent retrieval of LDAP objects 1386
LDAP sample files 1387
Installing the schema definition on the LDAP server 1388
Using the sample LDAP objects 1388
Defining QoS policies using LDAP 1389
Differentiated Services policy example 1389
RSVP policy example 1393
Sysplex distributor routing policy example 1395
Defining IDS policies using LDAP 1397
IDS scan policy example 1397
IDS attack policy example 1399
IDS TCP traffic regulation policy example 1406
IDS UDP traffic regulation policy example 1408

Appendix G. Related protocol specifications 1409

Appendix H. Accessibility 1429

Notices 1433
Terms and conditions for product documentation 1434
IBM Online Privacy Statement 1435
Policy for unsupported hardware 1435
Minimum supported hardware 1435
Policy for unsupported hardware 1436
Trademarks 1436

Bibliography 1437

Index 1441

Communicating your comments to IBM 1457

Figures

1. z/OS Communications Server TCP/IP protocol suite....4

2. syslogd operation....30

3. Generic server....46

4. Server with affinity for a specific transport provider....47

5. Example of binding an application to a specific transport provider....48

6. REXX program to switch TSO user to another TCP/IP stack....53

7. SYS1.PARMLIB(BPXPRMxx) for CINET....54

8. Primary router per VLAN (shared OSA with multiple primary routers)....63

9. Single OSA and VLAN switch configuration....65

10. Matching VLAN switch configuration to multiple OSAs (VLAN configuration)....66

11. Single stack using multiple OSAs on the same physical network....67

12. Example IPv4 IPAQIDIO profile before conversion....77

13. Example IPv4 IPAQIDIO profile after conversion....77

14. HiperSockets internal LAN....78

15. HiperSockets multiple internal LANs....79

16. Spanned IQD (HiperSockets) CHPID....81

17. Candidate configuration for HiperSockets Accelerator....85

18. HiperSockets Accelerator configuration....86

19. HiperSockets Converged Interface overview....90

20. Managing HiperSockets networks without HSCI....92

21. Managing HiperSockets networks with HSCI....93

22. An HSCI sample configuration with two OSA-Express features, two IP interfaces and a single subnet....94

23. An HSCI sample configuration with a single OSA-Express feature, two IP interfaces and two subnets 95

24. Linux and z/VM VSwitch bridge environment 96

25. HCD IQD configuration panel (channel parameters) 100

26. Elements of a secure TCP/IP deployment 123

27. Stack access control overview 137

28. Port access control overview 138

29. Network access control example 142

30. IP filtering at the z/OS communication endpoint 150

31. Security protocols from a protocol layering perspective 153

32. e-business scenarios with virtual private networks 154

33. TN3270E Telnet server security overview 158

34. Using multiple Telnet ports to separate secure and non-secure traffic 158

35. Combining Telnet security with IPSec client-to-firewall authentication 159

36. Secure and non-secure traffic using a single Telnet port 160

37. FTP client and server TLS overview 160

38. zERT connection detail (SMF 119 subtype 11) record layout 165

39. zERT summary (SMF 119 subtype 12) record layout 170

40. Intrusion detection services overview 184

41. Defensive filtering overview 185

42. IKE daemon acting as an NSS client for a single TCP/IP stack 188

43. IKE daemon acting as an NSS client for multiple TCP/IP stacks 189

44. Example of TCP/IP operating characteristics in PROFILE.TCPIP 240

45. Example of physical characteristics in PROFILE.TCPIP 254

46. Example of reserved port number definitions 261

47. Syntax for TCP/IP message IDs 277

48. Sample network part 1 282
49. Sample network part 2, IPv6 OSPF AS 283
50. Static VIPA configuration 366
51. Examples of the profile before and after conversion 367
52. Sample DVIPA addressing in a sysplex environment 370
53. DVIPA takeover with SWSA 397
54. Sysplex distributor with SWSA 398
55. An example of TCP/IP and VTAM subplexes 437
56. z/OS multi-tier application load balancing using sysplex distributor and the OPTLOCAL keyword 487
57. Enhanced z/OS multi-tier application load balancing using sysplex distributor 489
58. z/OS multi-tier application configuration using CPCSCOPE DVIPAs 491
59. z/OS Communications Server IEDN-enabled HiperSockets overview 496
60. Shared Memory Communications over RDMA (SMC-R) 504
61. 10 GbE RoCE Express feature in a dedicated RoCE environment 505
62. 10 GbE RoCE Express feature in a shared RoCE environment 506
63. Shared Memory Communications - Direct Memory Access (SMC-D) 507
64. Rendezvous processing 508
65. Identifying an SMC-R link 511
66. Multiple TCP connections over one SMC-R link 511
67. SMC-R link group 512
68. Identifying an SMC-D link 513
69. RMBs assigned to an SMC-R link 514
70. Physical networks 520
71. Redundant SMC-R links in an SMC-R link group 522
72. Misleading full redundancy configuration in a shared RoCE environment 523

73. Failover processing within an SMC-R link group....524
74. Partially redundant SMC-R links....525
75. SMC-R link group with no redundant link....525
76. SMCv2 EID concepts....536
77. SMC-Dv2 and ISMv2 with PNetID....542
78. SMC-Dv2 and ISMv2 without PNetIDs....544
79. SMC-Dv2 and ISMv2 With and Without PNetIDs....546
80. SMC-Rv2 connectivity concepts....548
81. SMC-Rv2 OSA and RoCE relationship....551
82. SMC-Rv2 link group....554
83. Sysplex distributor and SMC-R interaction....556
84. Telnet connectivity....569
85. Telnet parameter placement....575
86. Telnet profiles and connections....580
87. Port 1023 connection characteristics....602
88. Mapping model....607
89. Search method....615
90. Session initiation failure scenarios....636
91. Session ending scenarios....637
92. Typical Telnet data flow....650
93. Time buckets....654
94. SSL/TLS-secured FTP session scenario 1....706
95. SSL/TLS-secured FTP session scenario 2....707
96. SSL/TLS-secured FTP session scenario 3....707
97. SSL/TLS-secured FTP session scenario 4....707

98. SSL/TLS-secured FTP session scenario 5 707
99. SSL/TLS-secured FTP session scenario 6 708
100. Hierarchical naming tree 730
101. Local caching-only name server example 750
102. Resolver caching example 751
103. Resolver caching process; each stack specifies one NSINTERRADDR value 753
104. Resolver caching process; each stack specifies multiple NSINTERRADDR values 754
105. Resolver related configuration files in z/OS UNIX and native MVS environments 772
106. Activating changes to your policies 792
107. Policy Agent roles 794
108. Sample policy infrastructure 799
109. Policy Agent configuration files 807
110. Using SECCLASS to identify interfaces 896
111. IPv4 packet encapsulated using AH in transport mode 922
112. IPv4 packet encapsulated using ESP in transport mode 922
113. IPv6 packet encapsulated using AH in transport mode 922
114. IPv6 packet encapsulated using ESP in transport mode 922
115. IPv4 packet encapsulated using AH in tunnel mode 923
116. IPv4 packet encapsulated using ESP in tunnel mode 923
117. IPv6 packet encapsulated using AH in tunnel mode 923
118. IPv6 packet encapsulated using ESP in tunnel mode 923
119. UDP-encapsulated transport mode 924
120. UDP-encapsulated tunnel mode 924
121. Symmetric encryption 925
122. Sample worksheet for stack security 936

123. Security model network..... 975
124. Trusted internal network model..... 976
125. Partner company model..... 985
126. Partner company with NAT model..... 999
127. Partner company with NAPT model..... 1012
128. Branch office model..... 1017
129. Branch office with NAT model..... 1026
130. Cascaded tunnels..... 1034
131. Nested tunnels..... 1035
132. z/OS host to z/OS host, double NAT..... 1037
133. z/OS host to non-z/OS host, double NAT..... 1038
134. z/OS in a host-to-security gateway configuration..... 1039
135. Enabling network security services..... 1044
136. NSS services by discipline..... 1084
137. NSS client authorization example..... 1093
138. Defensive filtering overview..... 1108
139. Application Transparent TLS..... 1121
140. Client application model..... 1141
141. Server application model..... 1142
142. Forked server application model..... 1143
143. z/OS Load Balancing Advisor..... 1148
144. Sample output for the MODIFY procname,DISPLAY,LB command..... 1177
145. Sample output for the MODIFY procname,DISPLAY,LB,INDEX=lbindex command, part 1 of 2..... 1179
146. Sample output for the MODIFY procname,DISPLAY,LB,INDEX=lbindex command, part 2 of 2..... 1180
147. Sample output for the MODIFY procname,DISPLAY,MEMBERS,DETAIL command..... 1185

148. z/OS Load Balancing Advisor configuration example....1188
149. System overview....1195
150. ADNR configuration example....1221
151. Overview of SNMP support....1242
152. Configuration files for SNMP agent....1245
153. Configuration files for snmp....1256
154. Configuration files for NetView SNMP....1258
155. Subagent connection to OSA/SF....1264
156. CSSMTP forwards mail messages from spool to the network....1283
157. sendmail to CSSMTP bridge....1298
158. Adding applications to /etc/inetd.conf....1327
159. Setting traces in /etc/inetd.conf....1327
160. Express Logon network....1335
161. Select processors....1339
162. Work with attached channel paths....1340
163. Initiate the Define Channel Path dialog....1341
164. Add channel path....1342
165. Specify Maximum Frame Size....1343
166. Define the channel path access list....1344
167. Channel path number FF defined....1345
168. Work with attached control units....1346
169. Add the control unit(s)....1347
170. Define a control unit....1348
171. Define it to the processor....1349
172. Currently defined control unit....1350

173. Define the devices 1351
174. Empty device list 1352
175. Define the devices for the control unit 1353
176. Add devices of type IQD 1354
177. Define number of devices 1355
178. Define device to operating system 1356
179. Select systems 1357
180. Complete the definition 1358
181. Definition completed 1359
182. Partial configuration for the IKE daemon and an NSS server 1369
183. Basic policy objects 1377
184. Complex policy conditions 1378
185. Complex policy conditions before explosion 1378
186. Rule-specific conditions and actions 1379
187. Reusable conditions and actions 1380
188. Policy groups 1380
189. LDAP schema object class hierarchy 1382

Tables

1. TCP/IP configuration data sets 15

2. Common environment variables 26

3. Environment variables by application 27

4. syslogd facilities 30

5. BPX.DAEMON 37

6. Program control 38

7. How your own socket programs select a stack 52

8. Communications Server CIM providers 110

9. User identification, authentication, and access control for z/OS Communications Server applications 124

10. SERVAUTH resource names used by TCP/IP 125

11. TCP/IP application load module and alias names 144

12. Socket option resource names 144

13. TCP/IP applications that set IPv6 advanced socket API options 145

14. Configuration data sets 151

15. Mode to use for different logging requirements 220

16. TCP/IP message catalogs 272

17. Interior Gateway Protocol characteristics 281

18. Multipath route limitations 296

19. Route precedence 323

20. Summary of dynamic VIPA creation results 405

21. DVIPA contention resolution for DVIPA definitions on the same stack 410

22. DVIPA contention resolution between stacks in the same CPC 411

23. DVIPA contention resolution between stacks in different CPCs..... 412

24. Weight determination..... 429

25. Sysplex problem monitoring.....457

26. Load balancing solution quick reference..... 472

27. Redundancy levels..... 526

28. Client mappings..... 616

29. Sliding-window round-trip average example.....653

30. Environment variables for z/OS UNIX Telnet..... 656

31. PORTCOMMAND scenarios.....671

32. Migrating existing FTP client configuration.....704

33. Migrating existing ciphers.....704

34. EZYFxxxx messages.....719

35. APIs that use resolver caching.....751

36. Local definitions available to resolver.....772

37. Policy components needed per policy type..... 788

38. Configuration files and policy definition files.....789

39. Policy formats..... 795

40. Configuration files used for various policy clients..... 813

41. Configuration files used for various policy clients..... 814

42. Where Policy Agent FLUSH and PURGE are configured.....827

43. How Policy Agent FLUSH and PURGE are used.....828

44. Effect of sensitivity level on the counting of scan events at each suspicion level.....852

45. Classification of ICMP events..... 853

46. Classification of ICMPv6 events..... 853

47. UDP port event classification.....854

48. TCP port event classification..... 854

49. Possible authentication and encryption combinations for a connection..... 921

50. Table of remote hosts and subnetworks..... 936

51. Expanded filter rule for internal traffic.....1033

52. Expanded filter rule for remote traffic..... 1033

53. Original and translated port values.....1065

54. SERVAUTH profile names for NSS.....1088

55. Mapped label names..... 1092

56. Interaction between the mode setting on the DmStackConfig statement and the mode setting in individual filters..... 1110

57. Comparison of IP security filters and defensive filters.....1112

58. AT-TLS configuration for the server system..... 1126

59. AT-TLS configuration for the client system..... 1128

60. ClientAuthType parameter settings..... 1133

61. Summary of selected Advisor display output fields and flags..... 1174

62. Summary of selected Agent display output flags.....1177

63. WLM WEIGHT - CP, zAAP, and zIIP fields..... 1183

64. Allowed quiesce and enable command sequences for members..... 1187

65. Dynamic updates of ADNR-managed zones by other entities..... 1211

66. Base sysplex and ADNR configuration..... 1216

67. ADNR application in a sysplex subplexing environment; Example 1.....1216

68. ADNR application in a sysplex subplexing environment; Example 2.....1217

69. ADNR application in a sysplex subplexing environment; Example 3.....1217

70. SNMPv1/SNMPv2c advantages and SNMPv3 disadvantages..... 1244

71. SNMPv1/SNMPv2c disadvantages and SNMPv3 advantages..... 1244

72. JES batch job and CSSMTP configuration combinations for secure mail..... 1294

73. Required AT-TLS policies when DCAS TLSMECHANISM is set to ATTLS..... 1319

74. Frame size specification..... 1343

About this document

This document contains guidance material to enable you to configure IP address spaces, servers, and applications for z/OS Communications Server. This volume is part of a two-volume set:

- z/OS Communications Server: IP Configuration Guide, which contains concepts and guidance, explaining an overall approach to IP configuration.
- z/OS Communications Server: IP Configuration Reference, which describes parameters, options, and syntax of statements.

The information in this document includes descriptions of support for both IPv4 and IPv6 networking protocols. Unless explicitly noted, descriptions of IP protocol support concern IPv4. IPv6 support is qualified within the text.

For detailed information about configuration-related data sets and statements, see z/OS Communications Server: IP Configuration Reference.

For detailed information about commands used during configuration, see z/OS Communications Server: IP System Administrator's Commands.

This document refers to Communications Server data sets by their default SMP/E distribution library name. Your installation might, however, have different names for these data sets where allowed by SMP/E, your installation personnel, or administration staff. For instance, this document refers to the *hlq*.SEZAINST samples data set simply as SEZAINST. Your installation might choose a data set name of SYS1.SEZAINST, CS390.SEZAINST, or other high level qualifiers for the data set name.

Who should read this document

This document is intended for programmers and system administrators who are familiar with TCP/IP, MVS™, z/OS UNIX, and the Time Sharing Option Extensions (TSO/E).

How this document is organized

This document is divided into the following parts:

Part 1, “Base TCP/IP system,” on page 1 contains information on configuring the base TCP/IP stack.

Part 2, “Server applications,” on page 567 contains information that explains the server applications for z/OS Communications Server.

Appendix A, “Setting up the InetD configuration file,” on page 1327 and other appendixes in part 3 provide extra information for this document.

“Notices” on page 1433 contains notices and trademarks that are used in this document.

“Bibliography” on page 1437 contains descriptions of the documents in the z/OS Communications Server library.

How to use this document

Use this document to perform the following tasks:

- Configure z/OS Communications Server
- Customize and administer z/OS Communications Server

How to contact IBM service

For immediate assistance, visit this website: https://www.ibm.com/mysupport

© Copyright IBM Corp. 2000, 2022

Most problems can be resolved at this website, where you can submit questions and problem reports electronically, and access a variety of diagnosis information.

For telephone assistance in problem diagnosis and resolution (in the United States or Puerto Rico), call the IBM Software Support Center anytime (1-800-IBM®-SERV). You will receive a return call within 8 business hours (Monday – Friday, 8:00 a.m. – 5:00 p.m., local customer time).

Outside the United States or Puerto Rico, contact your local IBM representative or your authorized IBM supplier.

If you would like to provide feedback on this publication, see "Communicating your comments to IBM" on page 1457.

Conventions and terminology that are used in this information

Commands in this information that can be used in both TSO and z/OS UNIX environments use the following conventions:

- When describing how to use the command in a TSO environment, the command is presented in uppercase (for example, NETSTAT).
- When describing how to use the command in a z/OS UNIX environment, the command is presented in bold lowercase (for example, **netstat**).
- When referring to the command in a general way in text, the command is presented with an initial capital letter (for example, Netstat).

All the exit routines described in this information are *installation-wide exit routines*. The installation-wide exit routines also called installation-wide exits, exit routines, and exits throughout this information.

The TPF logon manager, although included with VTAM®, is an application program; therefore, the logon manager is documented separately from VTAM.

Samples used in this information might not be updated for each release. Evaluate a sample carefully before applying it to your system.

z/OS no longer supports mounting HFS data sets (The POSIX style file system). Instead, a z/OS File System (ZFS) can be implemented. The term hierarchical file system, abbreviated as HFS, is defined as a data structure that has a hierarchical nature with directories and files. References to hierarchical file systems or HFS might still be in use in z/OS Communications Server publications.

Note: In this information, you might see the following Shared Memory Communications over Remote Direct Memory Access (SMC-R) terminology:

- RoCE Express®, which is a generic term representing IBM 10 GbE RoCE Express, IBM 10 GbE RoCE Express2, and IBM 25 GbE RoCE Express2 feature capabilities. When this term is used in this information, the processing being described applies to all of these features. If processing is applicable to only one feature, the full terminology, for instance, IBM 10 GbE RoCE Express will be used.
- RoCE Express2, which is a generic term representing an IBM RoCE Express2® feature that might operate in either 10 GbE or 25 GbE link speed. When this term is used in this information, the processing being described applies to either link speed. If processing is applicable to only one link speed, the full terminology, for instance, IBM 25 GbE RoCE Express2 will be used.
- RDMA network interface card (RNIC), which is used to refer to the IBM 10 GbE RoCE Express, IBM® 10 GbE RoCE Express2, or IBM 25 GbE RoCE Express2 feature.
- Shared RoCE environment, which means that the "RoCE Express" feature can be used concurrently, or shared, by multiple operating system instances. The feature is considered to operate in a shared RoCE environment even if you use it with a single operating system instance.

Clarification of notes

Information traditionally qualified as Notes is further qualified as follows:

Attention
Indicate the possibility of damage

Guideline
Customary way to perform a procedure

Note
Supplemental detail

Rule
Something you must do; limitations on your actions

Restriction
Indicates certain conditions are not supported; limitations on a product or facility

Requirement
Dependencies, prerequisites

Result
Indicates the outcome

Tip
Offers shortcuts or alternative ways of performing an action; a hint

How to read a syntax diagram

This syntax information applies to all commands and statements that do not have their own syntax described elsewhere.

The syntax diagram shows you how to specify a command so that the operating system can correctly interpret what you type. Read the syntax diagram from left to right and from top to bottom, following the horizontal line (the main path).

Symbols and punctuation

The following symbols are used in syntax diagrams:

Symbol
Description

►►
Marks the beginning of the command syntax.

►
Indicates that the command syntax is continued.

|
Marks the beginning and end of a fragment or part of the command syntax.

►◄
Marks the end of the command syntax.

You must include all punctuation such as colons, semicolons, commas, quotation marks, and minus signs that are shown in the syntax diagram.

Commands

Commands that can be used in both TSO and z/OS UNIX environments use the following conventions in syntax diagrams:

- When describing how to use the command in a TSO environment, the command is presented in uppercase (for example, NETSTAT).
- When describing how to use the command in a z/OS UNIX environment, the command is presented in bold lowercase (for example, netstat).

Parameters

The following types of parameters are used in syntax diagrams.

Required
Required parameters are displayed on the main path.

Optional
Optional parameters are displayed below the main path.

Default
Default parameters are displayed above the main path.

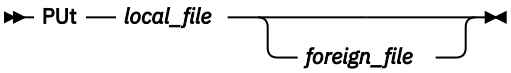
Parameters are classified as keywords or variables. For the TSO and MVS console commands, the keywords are not case sensitive. You can code them in uppercase or lowercase. If the keyword appears in the syntax diagram in both uppercase and lowercase, the uppercase portion is the abbreviation for the keyword (for example, OPERand).

For the z/OS UNIX commands, the keywords must be entered in the case indicated in the syntax diagram.

Variables are italicized, appear in lowercase letters, and represent names or values you supply. For example, a data set is a variable.

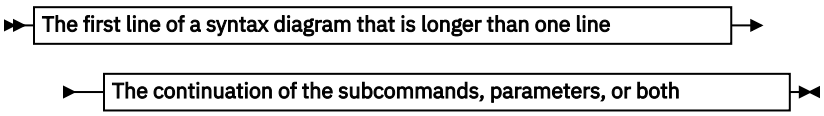
Syntax examples

In the following example, the PUt subcommand is a keyword. The required variable parameter is *local_file*, and the optional variable parameter is *foreign_file*. Replace the variable parameters with your own values.



Longer than one line

If a diagram is longer than one line, the first line ends with a single arrowhead and the second line begins with a single arrowhead.



Required operands

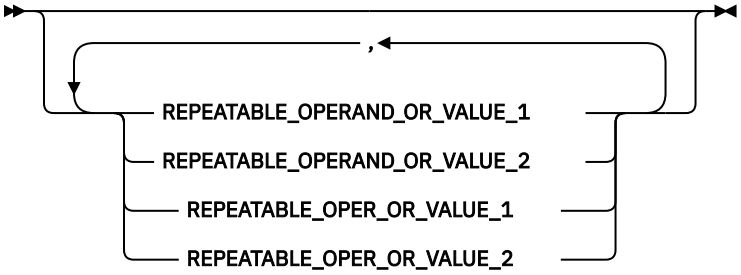
Required operands and values appear on the main path line. You must code required operands and values.

Optional values

Optional operands and values appear below the main path line. You do not have to code optional operands and values.

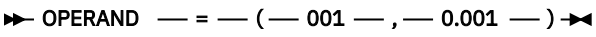
Selecting more than one operand

An arrow returning to the left above a group of operands or values means more than one can be selected, or a single one can be repeated.



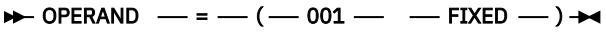
Nonalphanumeric characters

If a diagram shows a character that is not alphanumeric (such as parentheses, periods, commas, and equal signs), you must code the character as part of the syntax. In this example, you must code OPERAND=(001,0.001).



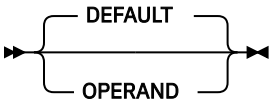
Blank spaces in syntax diagrams

If a diagram shows a blank space, you must code the blank space as part of the syntax. In this example, you must code OPERAND=(001 FIXED).



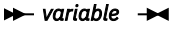
Default operands

Default operands and values appear above the main path line. TCP/IP uses the default if you omit the operand entirely.



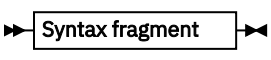
Variables

A word in all lowercase italics is a *variable*. Where you see a variable in the syntax, you must replace it with one of its allowable names or values, as defined in the text.



Syntax fragments

Some diagrams contain syntax fragments, which serve to break up diagrams that are too long, too complex, or too repetitious. Syntax fragment names are in mixed case and are shown in the diagram and in the heading of the fragment. The fragment is placed below the main diagram.



Syntax fragment

How to read a syntax diagram

This syntax information applies to all commands and statements that do not have their own syntax described elsewhere.

The syntax diagram shows you how to specify a command so that the operating system can correctly interpret what you type. Read the syntax diagram from left to right and from top to bottom, following the horizontal line (the main path).

Symbols and punctuation

The following symbols are used in syntax diagrams:

Symbol
Description

►►
Marks the beginning of the command syntax.

►
Indicates that the command syntax is continued.

|
Marks the beginning and end of a fragment or part of the command syntax.

►◄
Marks the end of the command syntax.

You must include all punctuation such as colons, semicolons, commas, quotation marks, and minus signs that are shown in the syntax diagram.

Commands

Commands that can be used in both TSO and z/OS UNIX environments use the following conventions in syntax diagrams:

- When describing how to use the command in a TSO environment, the command is presented in uppercase (for example, NETSTAT).
- When describing how to use the command in a z/OS UNIX environment, the command is presented in bold lowercase (for example, netstat).

Parameters

The following types of parameters are used in syntax diagrams.

Required
Required parameters are displayed on the main path.

Optional
Optional parameters are displayed below the main path.

Default
Default parameters are displayed above the main path.

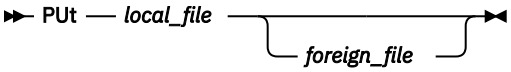
Parameters are classified as keywords or variables. For the TSO and MVS console commands, the keywords are not case sensitive. You can code them in uppercase or lowercase. If the keyword appears in the syntax diagram in both uppercase and lowercase, the uppercase portion is the abbreviation for the keyword (for example, OPERand).

For the z/OS UNIX commands, the keywords must be entered in the case indicated in the syntax diagram.

Variables are italicized, appear in lowercase letters, and represent names or values you supply. For example, a data set is a variable.

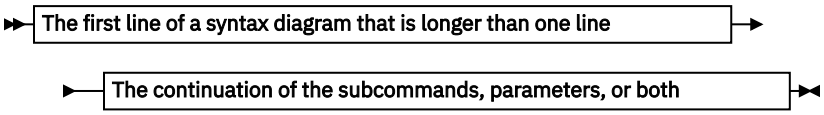
Syntax examples

In the following example, the PUt subcommand is a keyword. The required variable parameter is *local_file*, and the optional variable parameter is *foreign_file*. Replace the variable parameters with your own values.



Longer than one line

If a diagram is longer than one line, the first line ends with a single arrowhead and the second line begins with a single arrowhead.



Required operands

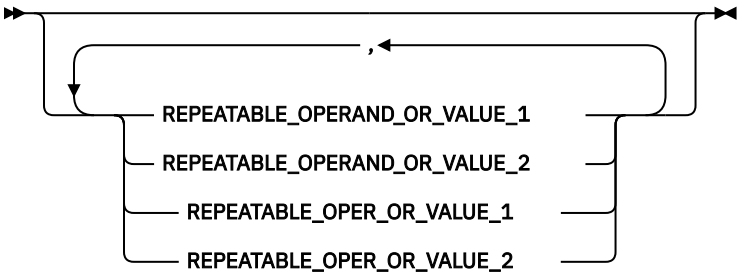
Required operands and values appear on the main path line. You must code required operands and values.

Optional values

Optional operands and values appear below the main path line. You do not have to code optional operands and values.

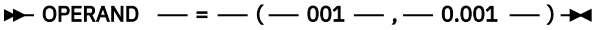
Selecting more than one operand

An arrow returning to the left above a group of operands or values means more than one can be selected, or a single one can be repeated.



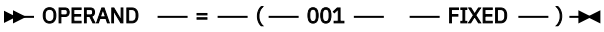
Nonalphanumeric characters

If a diagram shows a character that is not alphanumeric (such as parentheses, periods, commas, and equal signs), you must code the character as part of the syntax. In this example, you must code OPERAND=(001,0.001).



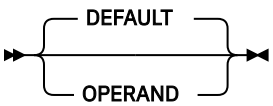
Blank spaces in syntax diagrams

If a diagram shows a blank space, you must code the blank space as part of the syntax. In this example, you must code OPERAND=(001 FIXED).



Default operands

Default operands and values appear above the main path line. TCP/IP uses the default if you omit the operand entirely.



Variables

A word in all lowercase italics is a *variable*. Where you see a variable in the syntax, you must replace it with one of its allowable names or values, as defined in the text.

Syntax fragments

Some diagrams contain syntax fragments, which serve to break up diagrams that are too long, too complex, or too repetitious. Syntax fragment names are in mixed case and are shown in the diagram and in the heading of the fragment. The fragment is placed below the main diagram.

Syntax fragment

Prerequisite and related information

z/OS Communications Server function is described in the z/OS Communications Server library. Descriptions of those documents are listed in “Bibliography” on page 1437, in the back of this document.

Required information

Before using this product, you should be familiar with TCP/IP, VTAM, MVS, and UNIX System Services.

Softcopy information

Softcopy publications are available in the following collection.

Titles	Description
IBM Z Redbooks	The IBM Z®® subject areas range from e-business application development and enablement to hardware, networking, Linux®, solutions, security, parallel sysplex, and many others. For more information about the Redbooks® publications, see http://www.redbooks.ibm.com/ and http://www.ibm.com/systems/z/os/zos/zfavorites/.

Other documents

This information explains how z/OS references information in other documents.

When possible, this information uses cross-document links that go directly to the topic in reference using shortened versions of the document title. For complete titles and order numbers of the documents for all products that are part of z/OS, see z/OS Information Roadmap (SA23-2299). The Roadmap describes what level of documents are supplied with each release of z/OS Communications Server, and also describes each z/OS publication.

To find the complete z/OS library, visit the z/OS library in IBM Documentation (https://www.ibm.com/docs/en/zos).

Relevant RFCs are listed in an appendix of the IP documents. Architectural specifications for the SNA protocol are listed in an appendix of the SNA documents.

The following table lists documents that might be helpful to readers.

Title	Number
DNS and BIND, Fifth Edition, O'Reilly Media, 2006	ISBN 13: 978-0596100575
Routing in the Internet, Second Edition, Christian Huitema (Prentice Hall 1999)	ISBN 13: 978-0130226471
sendmail, Fourth Edition, Bryan Costales, Claus Assmann, George Jansen, and Gregory Shapiro, O'Reilly Media, 2007	ISBN 13: 978-0596510299
SNA Formats	GA27-3136
TCP/IP Illustrated, Volume 1: The Protocols, W. Richard Stevens, Addison-Wesley Professional, 1994	ISBN 13: 978-0201633467
TCP/IP Illustrated, Volume 2: The Implementation, Gary R. Wright and W. Richard Stevens, Addison-Wesley Professional, 1995	ISBN 13: 978-0201633542
TCP/IP Illustrated, Volume 3: TCP for Transactions, HTTP, NNTP, and the UNIX Domain Protocols, W. Richard Stevens, Addison-Wesley Professional, 1996	ISBN 13: 978-0201634952
TCP/IP Tutorial and Technical Overview	GG24-3376
Understanding LDAP	SG24-4986
z/OS Cryptographic Services System SSL Programming	*SC14-7495*
z/OS IBM Tivoli Directory Server Administration and Use for z/OS	*SC23-6788*
z/OS JES2 Initialization and Tuning Guide	*SA32-0991*
z/OS Problem Management	*SC23-6844*
z/OS MVS Diagnosis: Reference	*GA32-0904*
z/OS MVS Diagnosis: Tools and Service Aids	*GA32-0905*
z/OS MVS Using the Subsystem Interface	*SA38-0679*
z/OS Program Directory	*GI11-9848*
z/OS UNIX System Services Command Reference	*SA23-2280*

Title	Number
z/OS UNIX System Services Planning	*GA32-0884*
z/OS UNIX System Services Programming: Assembler Callable Services Reference	*SA23-2281*
z/OS UNIX System Services User's Guide	*SA23-2279*
z/OS XL C/C++ Runtime Library Reference	*SC14-7314*
Open Systems Adapter-Express Customer's Guide and Reference	*SA22-7935*

Redbooks publications

The following Redbooks publications might help you as you implement z/OS Communications Server.

Title	Number
IBM z/OS Communications Server TCP/IP Implementation, Volume 1: Base Functions, Connectivity, and Routing	SG24-8096
IBM z/OS Communications Server TCP/IP Implementation, Volume 2: Standard Applications	SG24-8097
IBM z/OS Communications Server TCP/IP Implementation, Volume 3: High Availability, Scalability, and Performance	SG24-8098
IBM z/OS Communications Server TCP/IP Implementation, Volume 4: Security and Policy-Based Networking	SG24-8099
IBM Communication Controller Migration Guide	SG24-6298
IP Network Design Guide	SG24-2580
Managing OS/390 TCP/IP with SNMP	SG24-5866
Migrating Subarea Networks to an IP Infrastructure Using Enterprise Extender	SG24-5957
SecureWay Communications Server for OS/390 V2R8 TCP/IP: Guide to Enhancements	SG24-5631
SNA and TCP/IP Integration	SG24-5291
TCP/IP in a Sysplex	SG24-5235
TCP/IP Tutorial and Technical Overview	GG24-3376
Threadsafe Considerations for CICS	SG24-6351

Where to find related information on the Internet

z/OS

This site provides information about z/OS Communications Server release availability, migration information, downloads, and links to information about z/OS technology

http://www.ibm.com/systems/z/os/zos/

z/OS Internet Library

Use this site to view and download z/OS Communications Server documentation

http://www.ibm.com/systems/z/os/zos/library/bkserv/

z/OS Communications Server product

The page contains z/OS Communications Server product introduction

https://www.ibm.com/products/zos-communications-server

IBM Communications Server product support

Use this site to submit and track problems and search the z/OS Communications Server knowledge base for Technotes, FAQs, white papers, and other z/OS Communications Server information

https://www.ibm.com/mysupport

IBM Communications Server performance information

This site contains links to the most recent Communications Server performance reports

http://www.ibm.com/support/docview.wss?uid=swg27005524

IBM Systems Center publications

Use this site to view and order Redbooks publications, Redpapers, and Technotes

http://www.redbooks.ibm.com/

z/OS Support Community

Search the z/OS Support Community Library for Techdocs (including Flashes, presentations, Technotes, FAQs, white papers, Customer Support Plans, and Skills Transfer information)

z/OS Support Community

Tivoli® NetView® for z/OS

Use this site to view and download product documentation about Tivoli NetView for z/OS

http://www.ibm.com/support/knowledgecenter/SSZJDU/welcome

RFCs

Search for and view Request for Comments documents in this section of the Internet Engineering Task Force website, with links to the RFC repository and the IETF Working Groups web page

http://www.ietf.org/rfc.html

Internet drafts

View Internet-Drafts, which are working documents of the Internet Engineering Task Force (IETF) and other groups, in this section of the Internet Engineering Task Force website

http://www.ietf.org/ID.html

Information about web addresses can also be found in information APAR II11334.

Note: Any pointers in this publication to websites are provided for convenience only and do not serve as an endorsement of these websites.

DNS websites

For more information about DNS, see the following USENET news groups and mailing addresses:

USENET news groups
comp.protocols.dns.bind

BIND mailing lists
https://lists.isc.org/mailman/listinfo

BIND Users

- Subscribe by sending mail to bind-users-request@isc.org.
- Submit questions or answers to this forum by sending mail to bind-users@isc.org.

BIND 9 Users (This list might not be maintained indefinitely.)

- Subscribe by sending mail to bind9-users-request@isc.org.
- Submit questions or answers to this forum by sending mail to bind9-users@isc.org.

The z/OS Basic Skills Information Center

The z/OS Basic Skills Information Center is a web-based information resource intended to help users learn the basic concepts of z/OS, the operating system that runs most of the IBM mainframe computers in use today. The Information Center is designed to introduce a new generation of Information Technology professionals to basic concepts and help them prepare for a career as a z/OS professional, such as a z/OS systems programmer.

Specifically, the z/OS Basic Skills Information Center is intended to achieve the following objectives:

- Provide basic education and information about z/OS without charge
- Shorten the time it takes for people to become productive on the mainframe
- Make it easier for new people to learn z/OS

To access the z/OS Basic Skills Information Center, open your web browser to the following website, which is available to all users (no login required): https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zos.zbasics/homepage.html?cp=zosbasics

Summary of changes for IP Configuration Guide

This document contains terminology, maintenance, and editorial changes, including changes to improve consistency and retrievability. Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change.

Changes made in z/OS Communications Server Version 2 Release 5

This document contains information previously presented in z/OS Communications Server: IP Configuration Guide, which supported z/OS Version 2 Release 4. The most recent updates are listed at the top of each section.

New information

- Resolver support for z/OS data set encryption of basic and large format data sets, see "Configuration data set security" on page 150.
- Shared Memory Communications multiple IP subnet support (SMCv2), see the following topics:
  - "Shared Memory Communications multiple IP subnet support (SMCv2)" on page 534
    - "SMC filters" on page 515
    - "SMC Enterprise ID" on page 534
    - "EID Format" on page 537
    - "User Defined EID (UEID)" on page 537
    - "Multiple EIDs" on page 537
    - "System EID (SEID)" on page 538
    - "Using UEID or SEID" on page 538
    - "SMC filters" on page 539
    - "Using SMC filters" on page 539
    - "Dynamic selection of SMC version and SMC type" on page 540
    - "SMC-Dv2" on page 540
    - "ISMv2" on page 540
    - "Enabling SMC-Dv2" on page 540
    - "SMC-Dv2 and ISM VCHID Association using PNetIDs" on page 541
    - "SMC-Rv2" on page 547
    - "RoCEv2" on page 547
    - "Enabling SMC-Rv2" on page 548
    - "OSA and RoCEv2 device association and physical connectivity" on page 549
    - "SMC-Rv1 migration considerations" on page 552
    - "Sharing or reusing resources: SMCRIPADDR and PFID" on page 552
    - "SMC-Rv2 resilience, high availability with RoCEv2 and IP routing" on page 552
    - "SMC-Rv2 link groups" on page 552
- zERT policy-based enforcement, see the following topics:
  - "Secure Shell (SSH)" on page 162
  - "How does zERT enforcement work?" on page 172
  - "Enabling zERT policy-based enforcement" on page 173
  - "Defining a zERT enforcement policy" on page 176

© Copyright IBM Corp. 2000, 2022

- "zERT policy-based enforcement (ZERT) policy" on page 805

Changed information

- Notification of availability of TCP/IP extended services, see "Verifying your configuration" on page 268.
- IPsec certificate reporting enhancements, see the following topics:
  - "Displaying IKE tunnel information with the ipsec command" on page 1067
  - "Displaying the quick start Security Associations" on page 970
- Ping output enhanced to provide microsecond precision, see the following topics:
  - "Verifying connections with Netstat, Ping, and Traceroute" on page 358
  - "Activating the quick start Security Association" on page 970
- Removal of support for NCA policy import, see the following topics:
  - "Steps for configuring the Policy Agent" on page 808
  - "Step 1: Configure general information" on page 808
  - "Step 6: Configure Policy Agent for import services" on page 818
  - "FLUSH and PURGE considerations" on page 827
  - "Policy Agent" on page 793
  - "Import services" on page 796
  - "Policy Agent services" on page 794
  - "Refreshing policies" on page 827
  - "Sample policy infrastructure" on page 798
  - "Add TLS/SSL to Policy Agent connections" on page 822
  - "Verifying that policies are correctly defined and functioning properly" on page 830
- Removal of support for load balancing to Data Power®, see the following topics:
  - "Configuring distributed DVIPAs - sysplex distributor" on page 382
  - "Route selection for distributing packets" on page 385
  - "Sysplex distributor" on page 475
  - "TIER1, TIER2, and CPCSCOPE keyword DVIPA contention resolution" on page 409
- Removal of support for HFS, see the following topics:
  - "Hierarchical file system concepts" on page 12
  - "Other considerations when starting the Policy Agent" on page 825
- Removal of native TLS/SSL support from TN3270E Telnet Server, FTP Server, and DCAS, see the following topics:
  - "What DCAS provides" on page 1317
  - "Customizing DCAS for TLS/SSL" on page 1318
  - "Multiple port support" on page 158
  - "Secure and non-secure connections using a single Telnet port" on page 159
  - "Transport Layer Security" on page 598
  - "Connection persistence" on page 604
  - "Steps for customizing the FTP server for TLS" on page 686
  - "Steps for migrating the FTP client to use AT-TLS" on page 703
- Removal of Sysplex Distributor support for Cisco Multi-Node Load Balancer (MNLB), see the following topics:
  - "Using Netstat support to verify dynamic VIPA configuration" on page 423
  - "Configuring VIPAs for activation with VIPABACKUP" on page 417

- zERT policy-based enforcement, see the following topics:
  - "Logging of system messages" on page 30
  - "zERT information service access control" on page 147
  - "Monitoring cryptographic network protection: z/OS encryption readiness technology (zERT)" on page 162
  - "z/OS Encryption Readiness Technology (zERT) Concepts" on page 163
  - "How does zERT discovery provide the information?" on page 164
  - "What are the limitations for zERT discovery?" on page 167
  - "What does zERT aggregation collect?" on page 168
  - "How does zERT aggregation provide the information?" on page 169
  - "Using z/OS Encryption Readiness Technology" on page 172
  - "Enabling zERT discovery" on page 172
  - "Selecting a destination for zERT discovery SMF records" on page 174
  - "Disabling zERT discovery" on page 175
  - "Policy types and infrastructure overview" on page 787
  - "Configuration files and policy definition files" on page 789
  - "Storing configuration files and policy definition files" on page 790
  - "TCP/IP stack" on page 793
  - "Policy Agent policies" on page 795
  - "Policy sample files" on page 800
  - "Policy configuration files" on page 806
  - "Step 1: Configure general information" on page 808
  - "Step 2: Configure Policy Agent as a policy server" on page 811
  - "Step 4: Configure policies in Policy Agent configuration files" on page 817
  - "Stopping the Policy Agent" on page 827
  - "FLUSH and PURGE considerations" on page 827
  - "Switching between local and remote policies" on page 829
  - "Traffic regulation management daemon" on page 797

Removed information

- Removal of support for load balancing to Data Power
  - Sysplex distribution with DataPower®
- Removal of native TLS/SSL support from TN3270E Telnet Server, FTP Server, and DCAS:
  - *Migrating the DCAS server to use AT-TLS policies*

    Note: This section is moved to z/OS Upgrade Workflow.
- Removal of Sysplex Distributor support for Cisco Multi-Node Load Balancer (MNLB), see the following topics:
  - *Optimized connection load balancing using sysplex distributor in a network with CISCO routers (IPv4 only)*
  - *Steps for setting up sysplex distributor to be the service manager for the Cisco MNLB (IPv4 only)*

Changes made in z/OS Communications Server Version 2 Release 4

This document contains information previously presented in z/OS Communications Server: IP Configuration Guide, which supported z/OS Version 2 Release 3. The most recent updates are listed at the top of each section.

New information

September 2020 refresh

- Resolver support for z/OS data set encryption of basic and large format data sets, see "Configuration data set security" on page 150.

August 2020 refresh

- Shared Memory Communications multiple IP subnet support (SMCv2), see the following topics:
  - "Shared Memory Communications multiple IP subnet support (SMCv2)" on page 534
    - "SMC Enterprise ID" on page 534
    - "EID Format" on page 537
    - "User Defined EID (UEID)" on page 537
    - "Multiple EIDs" on page 537
    - "System EID (SEID)" on page 538
    - "Using UEID or SEID" on page 538
    - "Enabling SMC-Dv2" on page 540
    - "SMC-Dv2 and ISM VCHID Association using PNetIDs" on page 541

Prior to August 2020 refresh

- zERT Aggregation recording interval (APAR PH25049), see "Defining the zERT aggregation recording interval" on page 173.
- IBM Health Checker for use of native TLS/SSL support for the FTP server (APAR PH21573 and OA59022), see the following topics:
  - "Steps for customizing the FTP server for TLS" on page 686
  - "Steps for migrating the FTP client to use AT-TLS" on page 703
- Sysplex Autonomics for IPSec (APAR PH12788), see "Sysplex autonomics for IPsec" on page 465.
- Network support for z/OS Container Extensions, see the following topics:
  - "IBM z/OS Container Extensions network overview" on page 110
  - "Configuring application-instance DVIPAs for IBM z/OS Container Extensions (zCX)" on page 380
- AT-TLS support for TLS v1.3, see "Using TLSv1.3 protocol support" on page 1135.

Changed information

August 2021 refresh

- Shared Memory Communications multiple IP subnet support (SMCv2), see the following topics:
  - "SMC-Dv2 and ISM VCHID Association using PNetIDs" on page 541
  - "Rendezvous processing" on page 507
  - "Enabling SMC-Dv2" on page 540

November 2020 refresh

"IBM z/OS Container Extensions network overview" on page 110

August 2020 refresh

- Shared Memory Communications multiple IP subnet support (SMCv2), see the following topics:
  - "Physical network considerations" on page 520

- "VLANID considerations" on page 518
- "Configuring Shared Memory Communications - Direct Memory Access" on page 532

Prior to August 2020 refresh

- zERT Aggregation recording interval (APAR PH25049), see the following topics:
  - "z/OS Encryption Readiness Technology (zERT) Concepts" on page 163
  - "How does zERT aggregation provide the information?" on page 169
  - "Enabling zERT aggregation" on page 173
  - "Setting up TCP/IP operating characteristics in PROFILE.TCPIP" on page 238
- Inbound Workload Queueing (IWQ) support for IBM z/OS Container Extensions (VTAM APAR OA58300 and TCP/IP APAR PH16581), see "QDIO inbound workload queueing" on page 71.
- SMTPD compatibility enhancements for CSSMTP (APAR PH18237), see "Customizing the CSSMTP configuration file to handle undeliverable mail" on page 1290.
- Sysplex Autonomics for IPSec (APAR PH12788), see the following topics:
  - "Sysplex problem detection and recovery" on page 452
  - "Problem detection" on page 453
  - "Recovery" on page 455
  - "Summary of problems monitored and actions taken" on page 457
- Communications Server resolver enhancements, see the following topics:
  - "The resolver setup file" on page 738
  - "Steps for creating a resolver setup file" on page 742
  - "Information that is cached by the resolver" on page 751
  - "Steps for configuring resolver caching (optional)" on page 757
  - "Steps for manually managing the storage capacity of the resolver cache" on page 759
- Network support for z/OS Container Extensions, see the following topics:
  - "Connectivity and gateway functions" on page 5
  - "Terminology" on page 361
  - "Introduction to VIPA" on page 361
  - "Configuring DVIPA support" on page 369
- AT-TLS support for TLS v1.3, see the following topics:
  - Chapter 3, "Security," on page 123
  - "Existing workload" on page 153
  - Sockets return codes (ERRNOs)
  - "TN3270E Telnet server security" on page 157
  - "Multiple port support" on page 158
  - "Secure and non-secure connections using a single Telnet port" on page 159
  - "Express Logon Feature" on page 160
  - "Application Transparent Transport Layer Security" on page 161
  - "Cipher suite specification" on page 1134
  - "Protocol versions" on page 1134
  - "Certificate validation" on page 1135
  - "Encryption key refresh" on page 1136
  - "Handshake timer" on page 1137
  - "Wireless performance" on page 1139

- "Certificate selection" on page 1139
- "Session caching" on page 1139
- "AT-TLS access control considerations" on page 1140
- "Client application model" on page 1140
- "Server application model" on page 1141
- "Forked server application model" on page 1142
- "AT-TLS controlling application considerations" on page 1144
- "What are the limitations for zERT discovery?" on page 167

Changes made in z/OS Communications Server Version 2 Release 3

This document contains information previously presented in z/OS Communications Server: IP Configuration Guide, which supported z/OS Version 2 Release 2.

July 2018

Changed information

- IWQ support for IPSec, see "QDIO inbound workload queueing" on page 71.
- Code page enhancements for CSSMTP, see "Steps for creating mail on the JES spool data set for CSSMTP" on page 1287.

March 2018

New information

- HiperSockets Converged Interface support, see the following topics:
  - "z/OS HiperSockets connectivity" on page 88
    - "HiperSockets Converged Interface overview" on page 88
    - "Linux and z/VM VSwitch bridge considerations" on page 95
    - "Performance considerations for HiperSockets Converged Interface" on page 97
    - "SMC and HSCI PNetID considerations" on page 98
    - "Steps for enabling HiperSockets Converged Interface" on page 99
- z/OS Encryption Readiness Technology (zERT) aggregation, see the following topics:
  - "What does zERT aggregation collect?" on page 168
  - "How does zERT aggregation summarize the information?" on page 168
  - "How does zERT aggregation provide the information?" on page 169
  - "How does zERT aggregation determine the server port?" on page 171
  - "Enabling zERT aggregation" on page 173
  - "Selecting a destination for zERT aggregation SMF records" on page 174
  - "Disabling zERT aggregation" on page 175

Changed information

- z/OS Encryption Readiness Technology (zERT) aggregation, see the following topics:
  - "z/OS Encryption Readiness Technology (zERT) Concepts" on page 163
  - "What are the limitations for zERT discovery?" on page 167
  - "Using z/OS Encryption Readiness Technology" on page 172
  - "Enabling zERT discovery" on page 172

  - “Selecting a destination for zERT discovery SMF records” on page 174
  - “Disabling zERT discovery” on page 175
- TN3270E Telnet server Express Logon Feature support for Multi-Factor Authentication, see the following topics:
  - “Express Logon Feature” on page 643
  - Chapter 30, “Express logon services with the Digital Certificate Access Server,” on page 1317
  - Appendix C, “Express Logon Feature,” on page 1335
    - “Configuring RACF services for Express Logon” on page 1336
    - “An example of configuring the Express Logon components” on page 1337

January 2018

Maintenance and terminology changes are made for z/OS Version 2 Release 3 in January 2018.

September 2017

New information

- Permit Communications Server components to ICSF resources required by Network Authentication Service (Kerberos), see “Considerations for Kerberos authentication” on page 1314.
- sendmail to CSSMTP bridge, see the following topics:
  - “sendmail to CSSMTP bridge” on page 1297
  - “Setting up the sendmail bridge” on page 1298
  - “Steps for creating and customizing the configuration file” on page 1298
  - “Security considerations for the sendmail bridge” on page 1299
- z/OS Encryption Readiness Technology, see the following topics:
  - “zERT information service access control” on page 147
  - “Monitoring cryptographic network protection: z/OS encryption readiness technology (zERT)” on page 162

Changed information

- AT-TLS currency with System SSL, see “FIPS 140-2 support” on page 1136.
- Communications Server support for 8 character TSO User IDs, see the following topics:
  - “Stopping OMPROUTE” on page 304
  - “Stopping the IKE daemon” on page 1049
  - “Stopping the NSS server” on page 1102
  - “Stopping the DMD” on page 1119
  - “Stopping the trap forwarder daemon” on page 1266
- Communications Server support for enhanced system symbols, see “MVS system symbols” on page 28.
- Communications Server support for RoCE Express2 feature, see the following topics:
  - Comparing 10 GbE RoCE Express feature environments
  - Dedicated RoCE environment
  - Shared RoCE environment
  - 10 GbE RoCE Express2 feature environment
  - Shared Memory Communications terms
  - VLANID considerations

  - Physical network considerations
  - SMC-R high availability considerations
  - System requirements for SMC-R in a dedicated RoCE environment
  - System requirements for SMC-R in a shared RoCE environment
  - Configuring Shared Memory Communications over RDMA
  - VTAM displays and tuning statistics
- CSSMTP customizable ATSIGN character for mail addresses, see "Steps for using Transport Layer Security for CSSMTP" on page 1293.
- Enhanced wildcard support for jobname on PORT and PORTRANGE statements, see the following topics:
  - "Controlling access to particular ports" on page 138
  - "Using the PORT statement to control access to all unreserved ports" on page 139
- Improved CSSMTP code page compatibility with target servers, see "Steps for creating mail on the JES spool data set for CSSMTP" on page 1287.
- Permit Communications Server components to ICSF resources required by Network Authentication Service (Kerberos), see the following topics:
  - "Kerberos" on page 662
  - "Steps for customizing the FTP server for Kerberos" on page 691
- Removal of support for legacy devices, see the following topics:
  - "Connectivity and gateway functions" on page 5
  - "Setting up physical characteristics in PROFILE.TCPIP" on page 250
  - "Devices that support ARP offload" on page 257
  - "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 305
- Shared Memory Communications - Direct Memory Access, see Chapter 10, "Shared Memory Communications," on page 503.
- sendmail to CSSMTP bridge, see "Steps for creating mail on the JES spool data set for CSSMTP" on page 1287.
- Sysplex-wide security associations (SWSA) scalability improvement, see "Loss of access to coupling facility" on page 399.
- z/OS Encryption Readiness Technology, see the following topics:
  - "Local user access control to TCP/IP resources using SAF" on page 125
  - "Setting up TCP/IP operating characteristics in PROFILE.TCPIP" on page 238

Deleted information

- Removal of SMTPD & sendmail, all references to SMTP and sendmail are deleted.
- Removal of support for legacy devices, the Token-ring multicast topic is deleted.

Part 1. Base TCP/IP system

© Copyright IBM Corp. 2000, 2022

Chapter 1. Overview of z/OS Communications Server

z/OS Communications Server provides a set of communications protocols that support peer-to-peer connectivity functions for both local and wide-area networks, including the most popular wide-area network, the Internet. z/OS Communications Server also provides performance enhancements that can benefit a variety of TCP/IP applications.

z/OS Communications Server provides both SNA and TCP/IP protocols for z/OS. The SNA protocols are provided by VTAM and include Subarea, Advanced Peer-to-Peer Networking, and High Performance Routing protocols. For more information about z/OS Communications Server SNA protocols, see z/OS Communications Server: SNA Network Implementation Guide.

Figure 1 on page 4 shows the z/OS Communications Server TCP/IP protocol suite (also called stack), whose functions include associated applications, transport- and network-protocol layers, and connectivity and gateway functions. z/OS Communications Server contains IPv6 support. See z/OS Communications Server: IPv6 Network and Application Design Guide for more detailed information.

The z/OS Communications Server protocol suite supports two TCP/IP environments:

- A native MVS environment in which users can exploit the popular TCP/IP protocols in MVS application environments such as batch jobs, started tasks, TSO, CICS® applications, and IMS applications.
- A z/OS UNIX System Services environment that lets you create and use applications that conform to the POSIX or XPG4 standard (a UNIX specification).

Note: z/OS Communications Server exploits z/OS UNIX services even for traditional MVS environments and applications. Prior to using TCP/IP services, therefore, a full-function mode z/OS UNIX environment—including a Data Facility Storage Management Subsystem (DFSMSdfp), a z/OS UNIX file system, and a security product (such as Resource Access Control Facility, or RACF®)—needs to be defined and active before z/OS Communications Server can be started successfully.

© Copyright IBM Corp. 2000, 2022

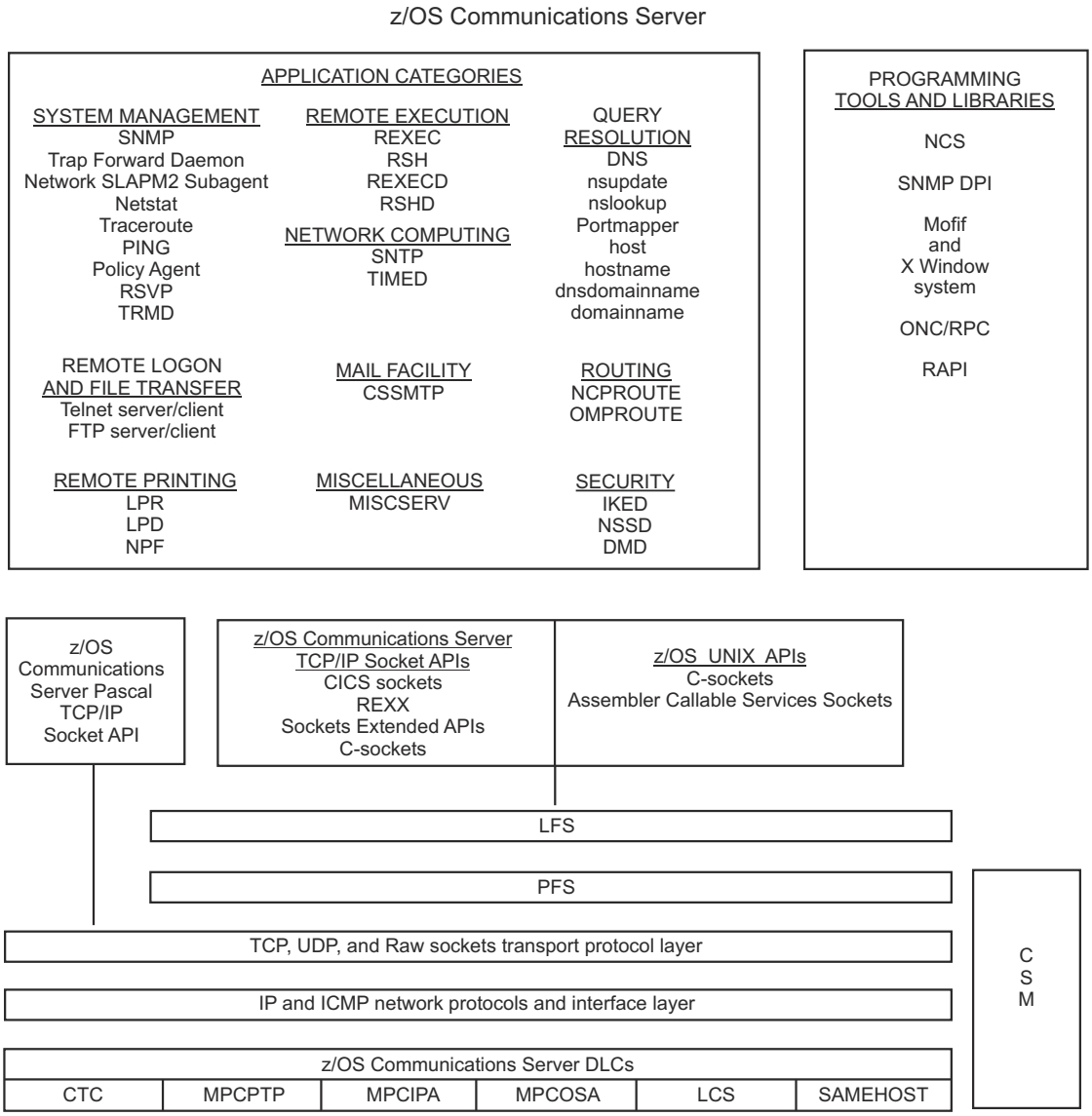


Figure 1. z/OS Communications Server TCP/IP protocol suite

The application categories in Figure 1 on page 4 do not list every application of the TCP/IP protocol suite.

z/OS Communications Server TCP/IP protocol-suite functions include the following categories:

- "TCP/IP protocol stack" on page 4
- "Connectivity and gateway functions" on page 5
- "Network protocol layer" on page 6
- "Transport layer" on page 7
- "File systems" on page 7
- "Application Programming Interfaces" on page 7

TCP/IP protocol stack

The Transmission Control Protocol (TCP) and the Internet Protocol (IP) refer to a non-proprietary protocol suite that enables different packet-switched networks to function as a single entity regardless of underlying network topology.

z/OS Communications Server provides robustness and high performance with the following features:

- A fully multiprocessor capable stack

- Exploitation of MVS Reliability, Availability, and Serviceability (RAS) services
- Exploitation of the z/OS architecture to optimize performance, CPU usage, and throughput
- Exploitation of the z/OS Sysplex functions to maximize availability and scalability of TCP/IP workloads

In addition, z/OS Communications Server design includes a tightly integrated storage and I/O model. I/O is provided by multipath channel (MPC) for network communication; storage management is provided by Communications Storage Manager (CSM).

Multipath channel I/O process

The term multipath channel (MPC) describes all z/OS Communications Server I/O driver support. There are specific I/O drivers under this support that are also referred to as MPC (such as MPCPTP).

MPC is a component of the I/O process model. MPC handles protocol headers and data separately and executes multiple I/O dispatchable units of work. MPC provides support for all devices supported by z/OS Communications Server. The MPC I/O process and the CSM facilities that TCP/IP exploits are part of the VTAM component of z/OS Communications Server. As a result, VTAM must be configured and active when starting devices on the TCP/IP stack.

Communications Storage Manager

The Communications Storage Manager (CSM) facility is used by authorized programs to manage subsystem storage pools. CSM reduces data moves by providing a flat storage model that is accessible at multiple layers of the process model and across MVS address space boundaries. In addition, CSM provides the following technical advantages:

- MVS cellpool-like services
- Automatic handling of the contraction and expansion of storage resources
- Handling of different types and sizes of storage requests (for example, pageable and fixed)

Connectivity and gateway functions

TCP/IP connectivity and gateway functions handle the physical interfaces and the routing of IP data packets called datagrams. The following communication interfaces are supported by z/OS Communications Server:

CTC
Provides access to TCP/IP hosts by way of a channel-to-channel (CTC) connection established over a System z ESCON channel.

LCS
Provides access to TCP/IP hosts using the following devices or features:

- An IBM 3172 Interconnect Controller, which connects to an Ethernet local area network
- An IBM 8232
- An IBM 2216 Multiaccess Connector Model 400
- An OSA-2 feature: Ethernet (EN)
- An OSA-Express feature: 1000BASE-T Ethernet or Fast Ethernet
- An OSA-Express2 feature: 1000BASE-T Ethernet

MPCIPA
Provides access to:

- TCP/IP hosts that use:
  - OSA-Express feature: Gigabit Ethernet (GbE), 1000BASE-T Ethernet, or Fast Ethernet
  - OSA-Express2 feature: GbE, 10 GbE, or 1000BASE-T Ethernet

  - OSA-Express3 or later feature: GbE or 10 GbE
  - HiperSockets: Uses the Internal Queued Direct I/O (iQDIO). HiperSockets provides high-speed, low-latency IP message passing between logical partitions (LPARs) within a single IBM eServer™ zSeries or later server.

  The OSA-Express features support the Queued Direct I/O (QDIO) architecture. IPv6 is supported on all Ethernet features in QDIO mode.
- An ensemble that uses:
  - OSA-Express3 or later feature configured as CHPID type OSX
  - OSA-Express3 or later feature configured as CHPID type OSM

MPCPTP
Provides access to TCP/IP hosts through multipath channel point-to-point (MPCPTP) links. MPCPTP supports IPv4 and IPv6 protocols. MPCPTP can be used in two ways to provide direct connectivity to other mainframe hosts running z/OS Communications Server:

- By using a set of two or more System z channels
- By configuring to use XCF services, if the System z hosts are part of the same sysplex

MPCPTP can also be used to provide the following connectivity options:

- Direct communication between two z/OS Communications Server TCP/IP Services protocol stacks running on the same MVS host without requiring any network attachments
- Connectivity to network attachments such as the IBM 2216 Multiaccess Controller Model 400 or the IBM RS/6000®

EZAZCX and EZ6ZCX
The EZAZCX (IPv4) and EZ6ZCX (IPv6) interfaces provide direct communications between the z/OS Communications Server TCP/IP protocol stack and instances of virtual servers executing in the IBM z/OS Container Extensions (zCX) address spaces. The interfaces are dynamically created and connected to the dynamically created VTAM TRLE called IUTZCX4n (IPv4) or IUTZCX6n (IPv6). The last character 'n' represents the instance of the zCX DLC created, where a unique instance of the zCX DLC is created for each TCP/IP stack providing services for zCX workloads.

IBM 10 GbE RoCE Express and RoCE Express2 feature
Enables the use of Remote Direct Memory Access (RDMA) processing by using Shared Memory Communications over RDMA (SMC-R) protocols.

The VTAM component of z/OS Communications Server provides the I/O support for each of these communication interfaces, and requires the creation (dynamically or through definition) of Transport Resource List entries (TRLEs) to represent each interface.

- TRLEs must be defined for the following communication interfaces:
  - MPCIPA devices not connected to an ensemble
  - MPCPTP

  For information about how to define these TRLEs, see z/OS Communications Server: SNA Resource Definition Reference.
- For all other communication interfaces, VTAM dynamically creates TRLEs. For information about dynamically created TRLEs for TCP/IP, see z/OS Communications Server: SNA Network Implementation Guide.

Network protocol layer

The network protocol layer provides the support for the IP protocol. All TCP and User Datagram Protocol (UDP) data goes through the IP layer when entering and leaving the host. TCP and UDP will use the IPv4 routing layer or the IPv6 routing layer.

The network layer also provides support for the Internet Control Message Protocol (ICMP) and ICMPv6. This is used by the IP layer to exchange information and error messages with IP layers on other hosts and routers. ICMP is used for the IPv4 protocol and ICMPv6 is used for the IPv6 protocol.

Transport layer

The transport layer provides the support for the TCP, UDP, and RAW protocols. All three protocols use IPv4 or IPv6 as the network layer. The TCP protocol provides a connection-oriented, reliable transport layer, whereas UDP provides a simpler, connectionless and unreliable transport layer. The RAW transport layer provides for a more direct interface to the IP layer, which is primarily used by system-management type applications.

File systems

The file system layer provides the main interface between the application programming interfaces (APIs) and the transport layers. The first component of the file system layer is the z/OS UNIX logical file system (LFS). The LFS provides the API layer with a common interface to access files and sockets. In a POSIX-compliant environment, applications can access both files and sockets in a similar fashion. For example, both files and sockets are represented by a 32-bit integer referred to as a descriptor. Common functions can be used to access both file and socket resources.

The layer beneath the LFS is the physical file system (PFS). The PFS layer is where the distinction between files, sockets, and other resources is made. Based on the resource type, the LFS passes the incoming function requests to the appropriate PFS, which handles requests related to resources in the z/OS UNIX file system. For more information about these physical file systems, see z/OS UNIX System Services Planning.

From a TCP/IP perspective, the AF_INET and the AF_INET6 PFS are of main interest. TCP/IP is enabled for IPv6 by defining an AF_INET6 PFS. Defining the file systems is the responsibility of the installation's z/OS UNIX programmer. The definitions are found in the BPXPRMxx member of SYS1.PARMLIB.

For information about defining AF_INET and AF_INET6 physical file systems, and about customizing BPXPRMxx for INET and CINET systems, see z/OS UNIX System Services Planning.

The AF_INET and the AF_INET6 PFS can be configured in two ways:

Integrated sockets PFS
The integrated sockets PFS can support the AF_INET PFS alone or AF_INET and AF_INET6 PFS together, but not AF_INET6 PFS alone.

Common INET PFS
This configuration is commonly referred to as the C_INET PFS configuration. It enables multiple AF_INET and AF_INET6 transport providers to be configured and active concurrently. Applications using the z/OS UNIX APIs do not know that multiple transport providers exist. For example, multiple TCP/IP Services components of z/OS Communications Server can be configured at the same time. The C_INET PFS is responsible for selecting the PFS over which to flow the request, based on the IP routing information from each of the AF_INET providers.

Under this configuration, it is also possible for TCP/IP application servers using the z/OS UNIX socket APIs to field incoming client requests from all AF_INET transport providers without knowing the particular transport provider.

Application Programming Interfaces

This information provides a short description of each of the application programming interfaces (APIs) that can be used to interface with the TCP/IP Services protocol stack provided by z/OS Communications Server. All of the APIs, with the exception of the PASCAL API, interface with the LFS layer.

The APIs are divided into the following two categories:

- TCP/IP socket APIs provided by z/OS Communications Server

- z/OS UNIX APIs

TCP/IP socket APIs provided by z/OS Communications Server

z/OS Communications Server provides several APIs to access TCP/IP sockets. These APIs can be used in either or both integrated and common INET PFS configurations. In a common INET PFS configuration, however, they function differently from z/OS UNIX APIs. In this type of configuration, the z/OS Communications Server APIs always bind to a single PFS transport provider, and the transport provider must be the TCP/IP stack provided by z/OS Communications Server.

The following TCP/IP socket APIs are included in z/OS Communications Server:

Pascal API
The Pascal application programming interface enables you to develop TCP/IP applications in Pascal language. Supported environments are normal MVS address spaces. The Pascal programming interface is based on Pascal procedures and functions that implement conceptually the same functions as the C socket interface. The Pascal routines, however, have different names than the C socket calls. Unlike the other APIs, the Pascal API does not interface directly with the LFS. It uses an internal interface to communicate with the TCP/IP protocol stack.

Pascal API supports only AF_INET.

CICS sockets
The CICS socket interface enables you to write CICS applications that act as clients or servers in a TCP/IP-based network. Applications can be written in C language, using the C sockets programming, or they can be written in COBOL, PL/I or assembly language, using the Sockets Extended programming interface.

CICS sockets support AF_INET and AF_INET6.

C sockets
The C sockets interface supports socket function calls that can be invoked from C programs. However, note that for C application development, IBM recommends the use of the UNIX C sockets interface. These programs can be ported between MVS and most UNIX environments relatively easily if the program does not use any other MVS specific services.

C sockets support only AF_INET.

IMS sockets
The Information Management System (IMS) IPv4 socket interface supports client/server applications in which one part of the application executes on a TCP/IP-connected host and the other part executes as an IMS application program.

IMS sockets support AF_INET and AF_INET6.

Sockets Extended macro API
The Sockets Extended macro API is a generalized assembler macro-based interface to sockets programming. It includes extensions to the socket programming interface, such as support for asynchronous processing on most sockets function calls.

The Sockets Extended macro API supports AF_INET and AF_INET6.

Sockets Extended Call Instruction API
The Sockets Extended Call Instruction API is a generalized call-based, high-level language interface to sockets programming. The functions implemented in this call interface resemble the C-sockets implementation, with some extensions similar to the sockets extended macro interface.

The Sockets Extended Call Instruction API supports AF_INET and AF_INET6.

REXX sockets
The REXX sockets programming interface implements facilities for socket communication directly from REXX programs by using an address rxsocket function. REXX socket programs can execute in TSO, online, or batch.

The REXX sockets programming interface supports AF_INET and AF_INET6.

See z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference for complete documentation of the TCP/IP Services APIs.

z/OS UNIX APIs

The following APIs are provided by the z/OS UNIX element of z/OS and are supported by the TCP/IP stack in z/OS Communications Server:

z/OS UNIX sockets

A set of C language functions provides support for the z/OS UNIX sockets, which are used in the z/OS UNIX environment. Programmers use this API to create applications that conform to the POSIX or XPG4 standard (a UNIX specification). Applications built with z/OS UNIX sockets can be ported to and from platforms that support these standards.

The z/OS UNIX sockets support AF_INET and AF_INET6.

z/OS UNIX assembler callable services

z/OS UNIX assembler callable services is a generalized call-based, high-level language interface to z/OS UNIX sockets programming. The functions implemented in this call interface resemble the z/OS UNIX C sockets implementation, with some extensions similar to the sockets extended macro interface.

The z/OS UNIX assembler callable services support AF_INET and AF_INET6.

For complete documentation of z/OS UNIX sockets, see z/OS XL C/C++ Runtime Library Reference. For information about z/OS UNIX callable services, see z/OS UNIX System Services Programming: Assembler Callable Services Reference.

Chapter 2. IP configuration overview

The objective of this topic is to help you be better prepared for installation related activities. It is important to understand the terms, relationships, and dependencies presented in this topic as a prerequisite to installation and customization. For additional background information on TCP/IP, see *TCP/IP Tutorial and Technical Overview*, GG24-3376, ISBN 0738421650.

IPv6 support

In this information, there might be no statement of support for a particular function in an IPv6 network. To determine whether a given function is applicable to IPv6, see the IPv6 support tables in z/OS Communications Server: IPv6 Network and Application Design Guide.

z/OS UNIX System Services concepts

An organization named X/Open documents standards of what to implement for UNIX interfaces in a series of guides that are published as the X/Open Portability Guides (XPG). X/Open owns the term UNIX and certifies different implementations of UNIX according to the UNIX definitions contained in XPG 4.2. z/OS UNIX System Services, or z/OS UNIX, is a certified UNIX system as defined by X/Open in XPG 4.2. z/OS UNIX coexists with traditional MVS functions and traditional MVS data set types such as partitioned data sets and sequential data sets. It enables access to z/OS UNIX files and utilities concurrently by using application programming interfaces (APIs) and the interactive shell environment. Two variants of the z/OS UNIX shell environment are available:

- The z/OS UNIX shell, much like a native UNIX environment
- The ISPF shell, an interface with access to menu-driven command interfaces

With the APIs, programs can run in any environment including batch jobs, in jobs submitted by TSO/E interactive users, and in most other started tasks, or in any other MVS application task environment. The programs can request:

- Only MVS services
- Only z/OS UNIX services
- Both MVS and z/OS UNIX services

The shell interface is an execution environment analogous to TSO/E, with a programming language of shell commands analogous to Restructured eXtended eXecutor (REXX) language. The shell support consists of:

- Programs that are run interactively by shell users
- Shell commands and scripts that are run interactively by shell users
- Shell commands and scripts that are run as batch jobs

Prior to OS/390® V2R5, OS/390 UNIX required APPC/MVS for programs issuing the fork() or spawn() function. APPC/MVS is no longer required for this purpose. Forked and spawned address spaces are implemented in z/OS for UNIX processing by the Work Load Manager (WLM) component of MVS.

- For a fork(), the system copies one process, called the parent process, into a new process, called the child process, and places the child process in a new address space, the forked address space.
- Spawn() also starts a new process in a new address space. Unlike a fork(), in a spawn() call the parent process specifies a name of a program to start the child process.

The types of processes can be:

- User processes, which are associated with a user
- Daemon processes, which perform continuous or periodic functions, such as a web server

© Copyright IBM Corp. 2000, 2022

Daemons are programs that are typically started when the operating system is initialized and remain active to perform standard services. Some programs that initialize processes for users are considered daemons, even though these daemons are not long-running processes. Examples of daemons provided by z/OS UNIX are cron, which starts applications at specific times, and inetd, which starts applications on demand.

A user or daemon process can have one or more threads. A thread is a single flow of control within a process. Application programmers create multiple threads to structure an application in independent sections that can run in parallel for more efficient use of system resources.

Overview of data sets and UNIX files

Data set and *file* are comparable terms. If you are familiar with MVS, you probably use the term *Data set* to describe a unit of data storage. If you are familiar with AIX® or UNIX, you probably use the term *file* to describe a named set of records stored or processed as a unit. In the TCP/IP environment, in addition to the traditional MVS data set organizations (such as sequential and partitioned), UNIX files are arranged in a hierarchical directory structure.

Some MVS data sets and UNIX files have special importance because of their function. For example, certain data sets and files are used when configuring the TCP/IP environment. Other data sets are used by the Telnet server when performing specific communication functions. For descriptions of the MVS data sets and UNIX files necessary for configuring the TCP/IP environment and the search orders used to find them, see Table 1 on page 15. A search order can include both MVS data sets and UNIX files, and these MVS data sets and UNIX files are collectively referred to as the *configuration files* in this information.

Note: Not all applications support UNIX files.

Hierarchical file system concepts

A hierarchical file system consists of:

- *Files*, which contain data or programs. A file containing a program object, shell script, or REXX program is called an *executable file*. Files are kept in directories.
- *Directories* that contain files, other directories, or both. Directories are arranged hierarchically, in a structure that resembles an upside down tree, with root directory at the top and the branches at the bottom. The *root* is the first directory for the file system at the peak of the tree and is designated by a slash (/).
- Named pipes, links, and other UNIX items, such as character special files like /dev/console that are used by applications like syslogd. See z/OS UNIX System Services Planning for more information about UNIX items like character special files.

The term *file system* has all of the following meanings:

- A logical collection of files, directories, named pipes, links, and other UNIX items and metadata that are arranged in a hierarchy.
- A particular instance of a logical collection of these items that are arranged in a hierarchy. They might be on local or remote disks or in computer memory.
- A program that is designed to provide the functions and data of one type of file system.

The context indicates which meaning is intended. Often more than one meaning is intended; this is an industry convention.

To the z/OS system, the file hierarchy is a collection of file systems. Additional instances of local or remote file systems might be mounted (logically connected) on directories of the root file system or of additional file systems.

Several types of file systems are supported by z/OS, including the following file systems:

- ZFS (z/OS File System)

  Each instance of z/OS File System is in a linear data set.

- TFS (temporary file system)

  Each instance of TFS is in computer memory.
- NFS (Network File System)

  NFS server provides access to file systems that are on other computers.

For most application programs, these types of file systems are interchangeable. The root file system is the first file system that is mounted. Subsequent file systems can be logically mounted on a directory within the root file system or on a directory within any mounted file system.

References to installation data sets

This information refers to Communications Server installation data sets by their default SMP/E distribution library name. Your installation might, however, have different names for these data sets where allowed by SMP/E, your installation personnel, or administration staff. For instance, this information refers to sample members of the *hlq*.SEZAINST library as SEZAINST(*member*), where the *hlq* value is the high-level qualifier specified during TCP/IP installation. Your installation can choose a data set name of SYS1.SEZAINST, CS390.SEZAINST, or other high level qualifiers for the data set name.

Understanding search orders of configuration information

It is important to understand the search order for configuration files used by TCP/IP functions, and when you can override the default search order with environment variables, JCL, or other variables you provide. This knowledge allows you to accommodate your local data set and file naming standards, and it is helpful to know the configuration data set or file in use when diagnosing problems.

It is important to note that the z/OS Communications Server environment consists of the TCP/IP address space, z/OS Communications Server applications, and the TCP/IP MVS applications. The TCP/IP address space functions are also referred to as the *stack*. The z/OS Communications Server applications refer to those applications using the z/OS UNIX socket API. The TCP/IP MVS applications refer to those applications written to the MVS APIs (for example, C, Sockets-Extended, CICS, IMS, and REXX). The TCP/IP stack and both sets of applications have some common (or global) configuration files, but they also use configuration files that are different.

Another important point to note is that when a search order is applied for any configuration file, the search ends with the first file found. Therefore, unexpected results are possible if you place configuration information in a file that never gets found, either because other files exist earlier in the search order, or because the file is not included in the search order chosen by the application.

Configuration data set naming conventions

When searching for configuration files, you can explicitly tell TCP/IP where most configuration files are by using DD statements in the JCL procedures or by setting environment variables. Otherwise, you can let TCP/IP dynamically determine the location of the configuration files, based on search orders shown in Table 1 on page 15.

For example, in Table 1 on page 15, for the FTP server application, if the installation did not code the //SYSFTPD DD statement, the FTP server would search for *jobname.*FTP.DATA, then file /etc/ftp.data, then data set SYS1.TCPPARMS(FTPDATA), and finally *hlq*.FTP.DATA.

Dynamic data set allocation

TCP/IP makes extensive use of dynamically allocated data sets using the MVS dynamic data set allocation function to search for configuration files. Multiple versions of a configuration data set can exist, each having a different high-level qualifier or middle-level qualifier. The search order for any configuration file will determine which data set is found and used.

High-level qualifier

High-level qualifiers (HLQ) permit you to associate an application's configuration data set with a particular job name or TSO user ID, or permit you to use a default configuration data set for the application. The possible high-level qualifiers are:

- *userid*

  Userid is the TSO user ID which invoked the application.

- *jobname*

  Jobname is the application's batch JCL job name or the name of the application's started procedure.

- *hlq*

  TCP/IP is distributed with a default high-level qualifier of TCPIP. To override the default HLQ used by dynamic data set allocation, specify the DATASETPREFIX statement in the TCPIP.DATA configuration file. For most configuration files, the DATASETPREFIX value is used as the high-level qualifier of the data set name in the last step in the search order. Note that the DATASETPREFIX value is not used as the high-level qualifier of the data set name used as the last step in the search order for the TCPIP.DATA configuration file.

Middle-level qualifiers

Multiple middle-level qualifiers (MLQ) permit the isolation of certain profile and translation table data sets. Two of the possible middle-level qualifiers are:

- Node name

  Node name is an MLQ used in the search order for finding the configuration file PROFILE.TCPIP. Node name is determined by the parameters specified during VMCF initialization. For further information on initializing VMCF, see *z/OS Program Directory*.

- Function name

  The TCP/IP implementation of multicultural support and double-byte character set (DBCS) support requires the use of multiple translation tables. To facilitate the concurrent use of multiple languages and code pages, TCP/IP uses a middle-level qualifier to designate which server or client uses a particular translation table. STANDARD, the default MLQ, is available for use if a single translation table can be used by multiple servers or clients. The TCP/IP Telnet client and FTP provide a TRANSLATE parameter that permits you to specify your chosen MLQ to replace the function name for that invocation of the command. For example, SRVRFTP is used as an MLQ by the File Transfer Protocol server.

The following data sets are some of the data sets that are only dynamically allocated by TCP/IP in a configuration file search order (you cannot specify them with DD statements in JCL):

```
ETC.PROTO               ETC.RPC
HOSTS.ADDRINFO          HOSTS.SITEINFO
SRVRFTP.TCPCHBIN        SRVRFTP.TCPHGBIN
SRVRFTP.TCPKJBIN        SRVRFTP.TCPSCBIN
SRVRFTP.TCPXLBIN        STANDARD.TCPCHBIN
STANDARD.TCPHGBIN       STANDARD.TCPKJBIN
STANDARD.TCPSCBIN       STANDARD.TCPXLBIN
```

For each of these data sets, the fully qualified name is established by using one of the following values as the data set HLQ:

- User ID or job name
- DATASETPREFIX value

Naming conventions for dynamically allocated data sets

A data set that you allocate explicitly (with a DD statement in JCL) can have any valid MVS data set name or z/OS UNIX file name. A data set that you create for the purpose of being allocated dynamically by TCP/IP must use the following naming conventions.

Note: In these examples, *xxxx* indicates an appropriate high-level qualifier, *yyyy* indicates an appropriate middle-level qualifier, and *zzzz* indicates an appropriate low-level qualifier.

- *userid.yyyy.zzzz*

  userid is the user ID of the logged on TSO user.

- *TSOprefix.yyyy.zzzz*

  TSOprefix is the data set prefix established by the TSO PROFILE command. *userid* is the default value of *TSOprefix*.

- *jobname.yyyy.zzzz*

  jobname is the job name specified on the JOB statement for a job stream or the procedure name for a started procedure.

- *hlq.yyyy.zzzz*

  hlq is the TCP/IP HLQ distributed as the system default, which can be overridden by the value in the DATASETPREFIX statement.

- *xxxx.nodename.zzzz*

  nodename is an MLQ that is used to define the data set name for the TCP/IP stack profile data set.

- *xxxx.function_name.zzzz*

  function_name denotes an acronym specifying a particular TCP/IP server (for example SRVRFTP for the FTP server) and is used as an MLQ for the translation table data set for that application.

- *xxxx.private_name.zzzz*

  private_name is a user-specified private qualifier that can be specified as an option on some TCP/IP commands.

- SYS1.TCPPARMS(TCPDATA)

  The member of a system data set used to find the *configuration file* TCPIP.DATA.

TCP/IP configuration data sets

Table 1 on page 15 lists the configuration MVS data sets and z/OS UNIX files used by the TCP/IP servers and functions. The table includes the name of the sample data set or file that is provided by Communications Server, and the way the data set or file is used.

Table 1. TCP/IP configuration data sets

Name (search order)	Copied from	Usage
ADNR.CONF The MVS data set or z/OS UNIX file specified on the CONFIG DD statement in the automated domain name registration started procedure	SEZAINST(ADNRCNF)	Contains automated domain name registration configuration statements.
CSSMTP.CONF 1. The MVS data set or z/OS UNIX file referenced by the CONFIG DD statement in the CSSMTP application started procedure 2. *jobname*.CSSMTP.CONF	SEZAINST(CSSMTPCF)	Contains CSSMTP application configuration statements.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
Defense Manager daemon (DMD) configuration 1. The MVS data set or z/OS UNIX file specified by the DMD_FILE environment variable 2. /etc/security/dmd.conf	/usr/lpp/tcpip/samples/dmd.conf	Contains DMD configuration statements.
Digital certificate access server (DCAS) configuration 1. The MVS data set or z/OS UNIX file that the DCAS_CONFIG_FILE environment variable specified 2. /etc/dcas.conf 3. *tsouserid*.DCAS.CONF 4. TCPIP.DCAS.CONF	No sample provided.	Contains DCAS configuration statements.
/etc/hosts	No sample provided.	One of the possible local host files used for IPv4 name query. See "Creating /etc/hosts" on page 267.
hlq.ETC.IPNODES	SEZAINST(EZBREIPN)	One of the local host files used for IPv6 name query, or IPv4 and IPv6 name query when COMMONSEARCH is specified in the resolver setup file.
/etc/mail/ezatmail.cf	/usr/lpp/tcpip/samples/ezatmail.cf	Defines sendmail to CSSMTP bridge configuration statements.
ETC.PROTO	usr/lpp/tcpip/samples/protocol	Used to map types of protocol to integer values to determine the availability of the specified protocol. Required by several z/OS Communications Server components. The search order depends on the type of application (z/OS UNIX or native MVS).
ETC.RPC	SEZAINST(ETCRPC)	Defines RPC applications to the Portmapper function.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
ETC.SERVICES	usr/lpp/tcpip/samples/services	Establishes port numbers for servers using TCP and UDP. Required for z/OS UNIX SNMP and OMPROUTE (if the RIP protocol is used). The search order depends on the type of application (z/OS UNIX or native MVS).
/etc/syslog.conf	/usr/lpp/tcpip/samples/syslog.conf	Configuration file for the syslog daemon (syslogd).
FTP.DATA 1. -f command line parameter (FTP client only) 2. The MVS data set or z/OS UNIX file specified on the SYSFTPD DD statement in the FTP server started procedure 3. *userid/jobname*.FTP.DATA 4. /etc/ftp.data 5. SYS1.TCPPARMS(FTPDATA) 6. *hlq*.FTP.DATA	SEZAINST(FTCDATA) for the client and (FTPSDATA) for the server	Overrides default FTP client and server parameters for the FTP server. For more information about the *hlq*, *jobname*, or *userid* values, see Chapter 12, "Transferring files using FTP," on page 663.
HOSTS.LOCAL	SEZAINST(HOSTS)	Input data set to MAKESITE for generation of HOSTS.ADDRINFO and HOSTS.SITEINFO.
IKE daemon configuration 1. The MVS data set or z/OS UNIX file specified by the IKED_FILE environment variable 2. /etc/security/iked.conf	/usr/lpp/tcpip/samples/iked.conf	Contains IKE configuration statements.
INETD.CONF The MVS data set or z/OS UNIX file specified on the EXEC DD statement in the INETD started procedure	/samples/inetd.conf	Provides configuration management statements of generic servers for the Internet Daemon (InetD). InetD handles rlogin, telnetd, rshd, rexec, and other applications. For more information about InetD, see z/OS UNIX System Services Planning.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
LBADV.CONF The MVS data set or z/OS UNIX file specified on the CONFIG DD statement in the z/OS Load Balancing Advisor started procedure	SEZAINST(LBADVCNF)	Contains z/OS Load Balancing Advisor configuration statements.
LBAGENT.CONF The MVS data set or z/OS UNIX file specified on the CONFIG DD statement in the z/OS Load Balancing Agent started procedure.	SEZAINST(LBAGECNF)	Contains z/OS Load Balancing Agent configuration statements.
LPD.CONFIG	SEZAINST(LPDDATA)	Configures the Line Printer Daemon for the Remote Print Server.
MIBS.DATA 1. The name of a z/OS UNIX file or an MVS data set specified by the MIBS_DATA environment variable 2. /etc/mibs.data z/OS UNIX file	No sample provided	Defines textual names for MIB objects for the z/OS UNIX snmp command.
Network security services (NSS) server configuration 1. The name of a z/OS UNIX file or MVS data set specified by the NSSD_FILE environment variable. 2. /etc/security/nssd.conf	/usr/lpp/tcpip/samples/nssd.conf	Contains NSS server configuration statements.
OMPROUTE configuration 1. The MVS data set or z/OS UNIX file specified on the OMPCFG DD statement in the OMPROUTE started procedure. 2. The MVS data set or z/OS UNIX file specified by the OMPROUTE_FILE environment variable 3. /etc/omproute.conf 4. *hlq*.ETC.OMPROUTE.CONF	SEZAINST(EZAORCFG)	Contains OMPROUTE configuration statements.
OSNMP.CONF 1. /etc/osnmp.conf 2. /etc/snmpv2.conf	/usr/lpp/tcpip/samples/snmpv2.conf	Defines target host security parameters for the osnmp command.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
OSNMPD.DATA 1. The MVS data set or z/OS UNIX file specified by the OSNMPD_DATA environment variable 2. /etc/osnmpd.data file system file 3. The MVS data set z/OS UNIX file specified on the OSNMPD DD statement in the agent started procedure 4. *jobname*.OSNMPD.DATA, where *jobname* is the name of the job used to start the SNMP agent 5. SYS1.TCPPARMS(OSNMPD) 6. *hlq*.OSNMPD.DATA, where *hlq* either defaults to TCPIP or is specified on the DATASETPREFIX statement in the TCPIP.DATA file being used	/usr/lpp/tcpip/samples/osnmpd.data	Used by SNMP for setting values for selected MIB objects.
PAGENT.CONF 1. File or data set specified with -c startup option 2. File or data set specified with PAGENT_CONFIG_FILE environment variable 3. /etc/pagent.conf	/usr/lpp/tcpip/samples/pagent.conf	Defines Policy Agent configuration parameters and optionally defines QoS service policies (rules and actions).
PROFILE.TCPIP 1. The MVS data set specified on the PROFILE DD statement in the TCP/IP stack started procedure. 2. *jobname.nodename*.TCPIP 3. TCPIP.*nodename*.TCPIP 4. *jobname*.PROFILE.TCPIP 5. TCPIP.PROFILE.TCPIP	SEZAINST(SAMPPROF)	Provides TCP/IP initialization parameters and specifications for network interfaces and routing.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
PW.SRC 1. The MVS data set or z/OS UNIX file specified by the PW_SRC environment variable 2. /etc/pw.src file system file 3. The MVS data set or z/OS UNIX file specified on SYSPWSRC DD statement in the started agent procedure 4. *jobname*.PW.SRC, where *jobname* is the name of the job used to start the SNMP agent 5. SYS1.TCPPARMS(PWSRC) 6. *hlq*.PW.SRC, where *hlq* either defaults to TCPIP or is specified on the DATASETPREFIX statement in the TCPIP.DATA file being used	No sample provided	Defines a list of community names used when accessing objects on a destination SNMP agent.
Resolver Setup File	SEZAINST (RESSETUP)	Provides configuration statements for the resolver.
RSVPD.CONF 1. File or data set specified with -c startup option 2. File or data set specified with RSVPD_CONFIG_FILE environment variable 3. /etc/rsvpd.conf 4. *hlq*.RSVPD.CONF	/usr/lpp/tcpip/samples/rsvpd.conf	Defines RSVP Agent configuration parameters.
SMTPNOTE clist System CLIST data set	SEZAINST(SMTPNOTE)	Defines the *note* parameters for the CSSMTP application.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
SNMPD.BOOTS 1. The name of a z/OS UNIX file system file or an MVS data set specified by the SNMPD_BOOTS environment variable. 2. /etc/snmpd.boots	No sample provided	Defines the SNMP agent security and notification destinations. **Note:** If the SNMPD.BOOTS file is not provided, the SNMP agent creates the file. If multiple SNMPv3 agents are running on the same MVS image, use the environment variable to specify different SNMPD.BOOTS files for the different agents. For security reasons, ensure unique engine IDs are used for different SNMP agents.
SNMPD.CONF 1. The name of a z/OS UNIX file system file or an MVS data set specified by the SNMPD_CONF environment variable. 2. /etc/snmpd.conf **Note:** The first file found in the search order is used.	/usr/lpp/tcpip/samples/snmpd.conf	Defines the SNMP agent security and notification destinations. **Note:** If the SNMPD.CONF file is found, the PW.SRC file and the SNMPTRAP.DEST files are not used.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
SNMPTRAP.DEST 1. The MVS data set or z/OS UNIX file specified by the SNMPTRAP_DEST environment variable 2. /etc/snmptrap.dest file system file 3. The MVS data set or z/OS UNIX file specified on SNMPTRAP DD statement in the agent started procedure 4. *jobname*.SNMPTRAP.DEST, where *jobname* is the name of the job used to start the SNMP agent 5. SYS1.TCPPARMS(SNMPTRAP) 6. *hlq*.SNMPTRAP.DEST, where *hlq* either defaults to TCPIP or is specified on the DATASETPREFIX statement in the TCPIP.DATA file being used	No sample provided	Defines a list of managers to which the SNMP agent sends traps.
TCPIP.DATA	SEZAINST(TCPDATA)	Provides parameters for TCP/IP client programs. The search order depends on the type of application (z/OS UNIX or native MVS). For more information about TCPIP.DATA, see z/OS Communications Server: IP Configuration Reference.
TNDBCSCN The MVS data set specified on the TNDBCSCN DD statement in the TN3270E Telnet server started procedure	SEZAINST(TNDBCSCN)	Provides configuration parameters for Telnet 3270 Transform support.

Table 1. TCP/IP configuration data sets (continued)

Name (search order)	Copied from	Usage
TRAPFWD.CONF 1. A z/OS UNIX system file or an MVS data set specified by the TRAPFWD_CONF environment variable 2. /etc/trapfwd.conf	No sample provided	Defines addresses to which the Trap Forwarder Daemon forwards traps. **Note:** If the environment variable is set and if the file specified by the environment variable is not found, the Trap Forwarder daemon terminates.
VTAMLST The VTAM definitions added to the ATCCONxx member of the MVS data set specified on the VTAMLST DD statement in the VTAM started procedure	SEZAINST(VTAMLST)	Defines VTAM applications and associated characteristics. Entries are required for the TN3270E Telnet server.

Configuration files for the TCP/IP stack

Two configuration files are used by the TCP/IP stack, PROFILE.TCPIP and TCPIP.DATA. PROFILE.TCPIP is used only for the configuration of the TCP/IP stack. TCPIP.DATA is used during configuration of both the TCP/IP stack and applications; the search order used to find TCPIP.DATA is the same for both the TCP/IP stack and applications.

PROFILE.TCPIP search order

During initialization of the TCP/IP stack, system operation and configuration parameters for the TCP/IP stack are read from the configuration file PROFILE.TCPIP. As shown in Table 1 on page 15, the search order used by the TCP/IP stack to find PROFILE.TCPIP involves both explicit and dynamic data set allocation as follows:

- //PROFILE DD DSN=*aaa.bbb.ccc(anyname)*
- *jobname.nodename*.TCPIP
- TCPIP.*nodename*.TCPIP
- *jobname*.PROFILE.TCPIP
- TCPIP.PROFILE.TCPIP

Note: Explicitly specifying the PROFILE DD statement in the TCPIPROC JCL is the recommended way to specify PROFILE.TCPIP. If this DD statement is present, the data set it defines is explicitly allocated by MVS and no dynamic allocation is done. If this statement is not present, the search order continues to use dynamic allocation for the PROFILE.TCPIP.

Examples

These examples show the search order used by TCP/IP to find the configuration file PROFILE.TCPIP. These examples use the sample TCP/IP started procedure, TCPIPROC, installed in the SEZAINST data set.

Example when DD cards are in your TCP/IP startup procedure

In this example, the PROFILE DD cards are specified as follows:

```
//TCPIP  PROC PARMS='CTRACE(CTIEZB00)'
//*
//* z/OS Communications Server
//* SMP/E Distribution Name: EZAEB01G
//*
//*     5694-A01 (C) Copr. IBM Corp. 1991,2001.
//*     US Government Users Restricted Rights -
//*     Use, duplication or disclosure restricted
//*     by GSA ADP Schedule Contract with IBM Corp.
//*     See IBM Copyright Instructions
//*
//TCPIP    EXEC PGM=EZBTCPIP,
//             PARM='&PARMS',
//             REGION=0K,TIME=1440
//*
⋮
//PROFILE DD DISP=SHR,DSN=MVSA.PROD.PARMS(PROFILE)
⋮
```

Because the PROFILE DD is the first step in the search order, TCP/IP uses the data set MVSA.PROD.PARMS(PROFILE) as the PROFILE.TCPIP configuration file.

Example when no DD cards are in your TCP/IP startup procedure

In this example, the PROFILE DD statement is not specified:

```
//TCPIP  PROC PARMS='CTRACE(CTIEZB00)'
//*
//* z/OS Communications Server
//* SMP/E Distribution Name: EZAEB01G
//*
//*     5694-A01 (C) Copr. IBM Corp. 1991,2001.
//*     US Government Users Restricted Rights -
//*     Use, duplication or disclosure restricted
//*     by GSA ADP Schedule Contract with IBM Corp.
//*     See IBM Copyright Instructions
//*
//TCPIP    EXEC PGM=EZBTCPIP,
//             PARM='&PARMS',
//             REGION=0K,TIME=1440
//*
⋮
```

For the configuration file PROFILE.TCPIP, the search order used is as follows:

1. PROFILE DD

   No PROFILE DD exists...search continues.

2. *jobname.nodename.*TCPIP

   If *jobname.nodename.*TCPIP is found, the search stops here.

3. TCPIP.*nodename*.TCPIP

   If TCPIP.*nodename*.TCPIP is found, the search stops here.

4. *jobname*.PROFILE.TCPIP

   If *jobname*.PROFILE.TCPIP is found, the search stops here.

5. TCPIP.PROFILE.TCPIP

   TCPIP.PROFILE.TCPIP is searched last if necessary.

TCPIP.DATA search order

The TCP/IP stack's configuration component uses the TCPIP.DATA configuration file during TCP/IP stack initialization to determine the stack's host name. To find the TCPIP.DATA information, the z/OS UNIX

environment search order is used. For a description of this search order, see “Search orders used in the z/OS UNIX environment” on page 777. This host name value is the value that is returned on gethostname socket function calls processed by this stack.

For details on the z/OS UNIX environment and native MVS environment search orders and the usage of z/OS UNIX environment variables, see “Resolver configuration files” on page 771.

Configuration files for TCP/IP applications

This information describes environment variables, the resolver configuration files that can be used by TCP/IP applications, and the search orders for those files. In addition to resolver files, an application can also have its own configuration files that are specific to that application. For more information about application-specific configuration files, see the descriptions of the individual applications in Part 2, “Server applications,” on page 567.

Environment variables

Environment variables are case-sensitive named variables with assigned values that various processes in the Communications Server configuration can access. Applications use environment variables to define the characteristics of their specific environment. Many z/OS Communications Server applications use environment variables to provide configuration information to the Language Environment® runtime functions, OMVS, the resolver, and to the application itself.

Use one of the following basic methods to specify environment variables:

- Specify the environment variable in the parameter field of the EXEC JCL statement in the started procedure.

  For example:

  ```
  //CONFHFS EXEC PGM=SYSLOGD,REGION=4096K,TIME=NOLIMIT,
  //        PARM='ENVAR("TZ=EST5EDT")/-c -i -f /user1/syslogd.conf'
  ```

  This method is useful when only one or two environment variables are required, as the maximum length of the PARM value is limited to 100 characters.
- Specify a file that contains the environment variables, which are processed during initialization of the application.

  For example:

  ```
  //CONFHFS EXEC PGM=SYSLOGD,REGION=4096K,TIME=NOLIMIT,
  //        PARM='ENVAR("_CEE_ENVFILE_S=DD:STDENV")/-c -i'
  ```

  This method is useful when multiple environment variables are required.

  The DD statement STDENV can point to a z/OS file system (zFS) file or to an MVS data set.

  – To allocate a zFS file on the STDENV DD statement:

    ```
    //STDENV   DD PATH='/etc/syslogd.env',PATHOPTS=(ORDONLY)
    ```

  – To allocate an MVS data set on the STDENV DD statement:

    ```
    //STDENV   DD DISP=SHR,DSN=TCPIP.SYSLOGD.ENV(SYSLOGD)
    ```

  When you are using the _CEE_ENVFILE_S environment variable, the data set organization can be a sequential data set or a member of a partitioned data set with a record format of Fixed, Fixed Blocked, Variable, or Variable Blocked. Trailing blank spaces are removed from the end of each record.

  Each record that contains an equal sign (=) is considered to be an environment variable. The text before the equal sign is the name of the environment variable. The text after the equal sign is the value of the environment variable. The case of the name and value are not changed.

You can use _CEE_ENVFILE_COMMENT as the first environment variable in the file to embed comments in the environment variable file. For example:

```
_CEE_ENVFILE_COMMENT=#
##############################################################
#   SYSLOGD environment variables
#   Last update =  2012/04/25
##############################################################
SYSLOGD_CODEPAGE=IBM_1047
SYSLOGD_DEBUG_LEVEL=0
```

The character that is specified for the _CEE_ENVFILE_COMMENT environment variable is then used in column 1 of a record to indicate that the record is a comment record and is to be ignored.

Rule: If you do not define the _CEE_ENVFILE_COMMENT environment variable, text strings of the form *name=value* are processed as environment variables. For example, if #SYSLOGD_CODEPAGE=IBM_1047 is encountered in an environment variable file, the environment variable #SYSLOGD_CODEPAGE is defined with the value IBM_1047.

For more information about using environment variables, see z/OS XL C/C++ Programming Guide.

Language Environment and UNIX System Services also provide environment variables. For information about these other variables, see Understanding Shell Variables in z/OS UNIX System Services User's Guide.

Table 2 on page 26 lists where to find more information about some common environment variables that z/OS Communications Server and its applications explicitly set or use.

Table 2. Common environment variables

Environment Variable	Description	Reference
_BPX_JOBNAME	Changes the job name when the application creates a new address space (FTPD)	_BPX environment variables in z/OS UNIX System Services Planning
_BPXK_SETIBMOPT_TRANSPORT	Requests transport affinity to a specific TCP/IP stack	*Requesting transport affinity* in z/OS UNIX System Services Planning
_CEE_ENVFILE	The name of an environment file, trailing blanks are left intact	_CEE_ENVFILE in z/OS XL C/C++ Programming Guide
_CEE_ENVFILE_S	The name of an environment file, trailing blanks are removed from each record	_CEE_ENVFILE_S in z/OS XL C/C++ Programming Guide
_CEE_ENVFILE_COMMENT	The character in column 1 of a record of the ENVFILE that indicates that the record is a comment record and is to be ignored	_CEE_ENVFILE_COMMENT in z/OS XL C/C++ Programming Guide
_CEE_ENVFILE_CONTINUATION	The last non-space character in a record that indicates that the environment variable value is continued on the next record	_CEE_ENVFILE_CONTINUATION in z/OS XL C/C++ Programming Guide
RESOLVER_CONFIG	The name of the TCPIP.DATA data set	Specifying TCP/IP address space parameters in z/OS Communications Server: IP Configuration Reference

Table 3 on page 27 lists where to find more information about the environment variables that particular z/OS Communications Server applications explicitly set or use.

Table 3. Environment variables by application	
Application	**For environment variable information, see:**
Communications Server SMTP (CSSMTP)	CSSMTP application environment variables in z/OS Communications Server: IP Configuration Reference
Defense Manager daemon (DMD)	“Steps for configuring the DMD” on page 1116
Digital certificate access server (DCAS)	DCAS environment variables in z/OS Communications Server: IP Configuration Reference
FTP client	Environment variables in z/OS Communications Server: IP User's Guide and Commands
FTP server	FTP server environment variables in z/OS Communications Server: IP Configuration Reference
IKE daemon	IKE environment variables in z/OS Communications Server: IP Configuration Reference
MIBDESC	MIBDESC environment variables in z/OS Communications Server: IP Configuration Reference
Motif	Motif environment variables in z/OS Communications Server: IP Programmer's Guide and Reference
Network security services (NSS) server	“Steps for configuring the NSS server” on page 1095
OMPROUTE	OMPROUTE environment variables in z/OS Communications Server: IP Configuration Reference
ORSHD	RSHD command (orshd) environment variables in z/OS Communications Server: IP Configuration Reference
OSNMP	OSNMP environment variables in z/OS Communications Server: IP Configuration Reference
Policy Agent	Policy Agent environment variables in z/OS Communications Server: IP Configuration Reference
Resolver	“z/OS XL C/C++ environment variables for configuration files” on page 775
SLAPM2 subagent (nslapm2)	Network SLAPM2 subagent environment variables in z/OS Communications Server: IP Configuration Reference
SNMP agent	OSNMPD environment variables in z/OS Communications Server: IP Configuration Reference
SNMP DPI	SNMP DPI environment variables in z/OS Communications Server: IP Programmer's Guide and Reference
TRAPFWD daemon	TRAPFWD environment variables in z/OS Communications Server: IP Configuration Reference
UNIX Telnet server (otelnetd)	“Environment variables” on page 656
X Window System interface V11R4 and Motif version 1.1	Environment variables in z/OS Communications Server: IP Programmer's Guide and Reference
X Window System	X Window System environment variables in z/OS Communications Server: IP Programmer's Guide and Reference

MVS-related considerations

This topic includes an overview of MVS-related considerations for configuring z/OS Communications Server.

MVS system symbols

Use of MVS system symbols in TCP/IP profile data sets is automatically supported. This automatic support first attempts to use hiperspace memory files to perform the symbol translation, but if an error occurs, a temporary file is used. The temporary file is created in the directory specified by the TMPDIR environment variable, or in the `/tmp` directory if the TMPDIR environment variable is not defined.

Use of MVS system symbols in the resolver setup file and the TCPIP.DATA file is also automatically supported. The resolver reads and processes the TCPIP.DATA file on behalf of TCP/IP applications that invoke resolver services. System symbols are resolved as file records are read.

If an MVS system symbol is used with a substitution value that has a length greater than the symbol name, and this substitution causes the record to overflow, then the symbol substitution will fail for the record.

Use of MVS system symbols is also supported in the following cases:

- Values of resolver environment variables, like RESOLVER_CONFIG and RESOLVER_TRACE
- OMPROUTE configuration file
- Communications Server SMTP (CSSMTP) configuration file
- BeginArchiveParms DSNPrefix parameter in the syslogd configuration file

For more information about the use of MVS system symbols, see What are system symbols in z/OS MVS Initialization and Tuning Reference.

EZACFSM1 symbol translator utility

For MVS system symbols in other configuration files, use the symbol translator utility, EZACFSM1, to translate the symbols before the files are read by TCP/IP. EZACFSM1 reads an input file and writes to an output file, translating any symbols in the process. The input file and output file can be MVS data sets or z/OS UNIX files; however, do not specify the same file for both the input and output file. For lists of the static system symbols and dynamic system symbols supported by EZACFSM1, see z/OS MVS Initialization and Tuning Reference.

EZACFMS1 returns the following error return codes:

- A return code of 45 if the same file was specified for both the input and the output.
- Errno value from z/OS UNIX fopen() function calls if errors occur opening the input or output file. See fopen() — Open a file in z/OS XL C/C++ Runtime Library Reference for a list of the return values for this function call.
- ASASYMBM error return codes. See ASASYMBM and ASASYMBF — Substitute text for symbols in z/OS MVS Programming: Assembler Services Reference ABE-HSP for a list of the return codes for this service.

The following symbol translator JCL is found in SEZAINST(CONVSYM) and is used to start EZACFSM1:

```
//________ JOB (accounting,information),programmer.name,
//             MSGLEVEL=(1,1),MSGCLASS=A,CLASS=A
//*
//*   CS for z/OS
//*   SMP/E distribution name: EZACFCSY
//*
//*   5694-A01 (C) Copyright IBM Corp. 1998, 2005
//*   Licensed Materials - Property of IBM
//*
//*   Function: System Symbols Translator JCL
//*
//*   This JCL kicks off a utility that will read from
//*   an input file that contains MVS System Symbols
//*   and produce an output file which has those symbols
```

```
//*   replaced with their substitution text, as defined
//*   in the appropriate IEASYMxx PARMLIB data set; see MVS
//*   Initializaton and Tuning Reference for rules about symbols.
//*
//*   This JCL can be run against any of the TCP/IP configuration
//*   files that contain MVS System Symbols.  An example of how it
//*   could be used is this; a customer could have one base TCPIP.DATA
//*   file containing MVS System Symbols which they edit and maintain.
//*   They would run this utility against this one file the various
//*   MVS systems to produce the TCPIP.DATA file for each different
//*   system.
//*
//STEP1   EXEC  PGM=EZACFSM1,REGION=0K
//SYSIN   DD DSN=TCP.DATA.INPUT,DISP=SHR
//*SYSIN   DD    PATH='/tmp/tcp.data.input'
//*          The input file can be either an MVS data set or an z/OS
//*            UNIX file.
//*
//*
//SYSOUT  DD DSN=TCP.DATA.OUTPUT,DISP=SHR
//*SYSOUT  DD    PATH='/tmp/tcp.data.output',PATHOPTS=(OWRONLY,OCREAT),
//*              PATHMODE=(SIRUSR,SIWUSR,SIRGRP,SIWGRP)
//*          The output file can be either an MVS data set or an z/OS
//*            UNIX file.
//*
//*          The output file cannot be the same file as the input file-
//*          doing so will result in a return code of 45.
//*
//*          You can mix input and output file types (i.e., the input
//*          can be an MVS data set with the output being a z/OS UNIX
//*          file or visa versa).
//*          Note: Other pathmodes for sysout may be used if needed.
```

The symbol translator utility can be used on any of the TCP/IP configuration files, but because the PROFILE.TCPIP file is automatically translated during TCP/IP initialization, there is no need to run the utility against that file.

Restriction: There is no input line length limitation. However, you must ensure that any symbols that require translation do not end on or span character position 80 or n*80 of the line. Otherwise, the utility cannot translate the symbol.

Automatic restart manager

Automatic restart manager (ARM) is an MVS component that can automatically restart the TCP/IP stack after an abnormal end (ABEND).

During initialization, TCP/IP automatically registers with the automatic restart manager using the following options:

```
REQUEST=REGISTER
ELEMNAME=EZAsysclonetcpname
ELEMTYPE=SYSTCPIP
TERMTYPE=ELEMTERM
```

where:

- *sysclone* is a 1– or 2–character shorthand notation for the name of the MVS system. For a complete description of the SYSCLONE static system symbol, see z/OS MVS Initialization and Tuning Reference.
- *tcpname* is a 1– to 8–character name of the TCP/IP stack which registers with the automatic restart manager. For example, if the SYSCLONE value is 02 and the TCP/IP stack name is TCPCS, the resulting ELEMENT value is EZA02TCPCS.

Rules:

- The TCP/IP stack supports automatic restart only on the same system. Do not explicitly specify TERMTYPE(ALLTERM) in the automatic restart management policy for TCP/IP. A policy that uses the default value ALLTERM is overridden by the TCP/IP stack registration value ELEMTERM.
- To prevent dynamic configuration changes made using the VARY TCPIP,,OBEYFILE command from being lost after restarting TCP/IP, ensure that these configuration changes are part of the normal TCP/IP startup procedure.

For more information about automatic restart manager, see z/OS MVS Setting Up a Sysplex.

Logging of system messages

Syslog daemon (syslogd) is a server process that must be started as one of the first processes in your z/OS UNIX environment. TCP/IP server applications and components use syslogd for logging purposes and can also send trace information to syslogd. Servers on the local system use AF_UNIX sockets to communicate with syslogd; remote servers use AF_INET sockets. z/OS Communications Server components use the `local1`, `local4`, `local5`, `daemon`, `mail`, `user`, and `auth` facilities names.

Note: Each application activates and deactivates traces in a slightly different manner. For details, see the information about the individual application.

The syslog daemon reads and logs system messages to the MVS console, log files, SMF, other machines, the operlog log stream, or users as specified by the configuration file. If syslogd is not started, log data from some applications will be displayed on the MVS console. For more information on syslogd, see Chapter 5, “TCP/IP Customization,” on page 213.

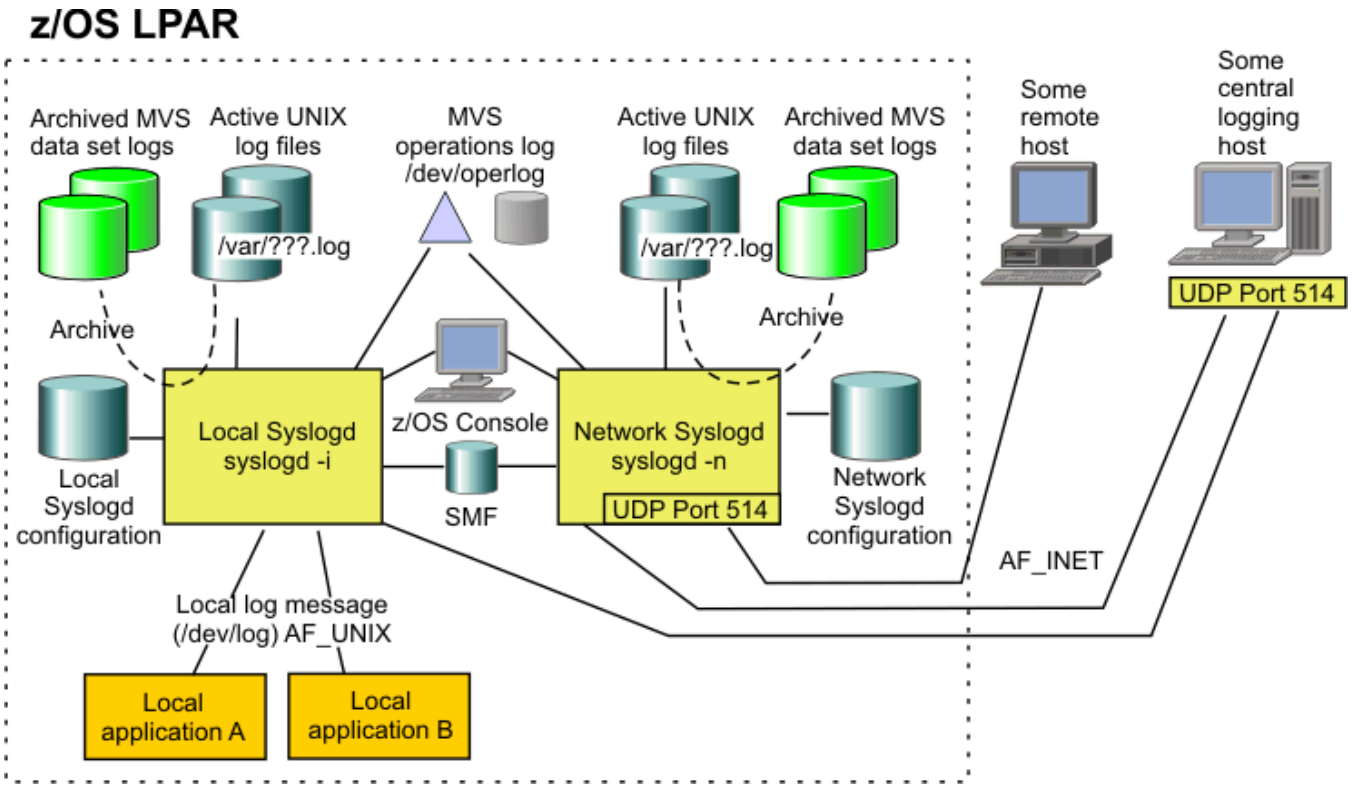


Figure 2. syslogd operation

Note: /var/???.log is the file specified in the syslogd.conf file.

The syslogd facility uses a common mechanism for separating messages. Table 4 on page 30 shows the facilities used by z/OS Communications Server functions which write messages to syslogd. The Primary syslog facility column shows the syslog facility used for most messages logged by the application. Some applications use other facilities for certain messages. Table 4 on page 30 also shows any additional facilities.

Table 4. syslogd facilities

Application	syslogd record identifications	Primary syslog facility	Other syslog facility
Application Transparent Transport Layer Security (AT-TLS)	TTLS	daemon	auth

Table 4. syslogd facilities (continued)

Application	syslogd record identifications	Primary syslog facility	Other syslog facility
Automated domain name registration (ADNR)	adnr	daemon	None
Communications Server SMTP (CSSMTP)	CSSMTP	mail	None
Defense Manager daemon (DMD)	DMD	local4	None
FTP server	ftpd, ftps	daemon	None
IKE daemon	IKED	local4	None
Network security services (NSS) server	NSSD	local4	None
Network SLAPM2 subagent	NSLAPM2	daemon	None
OMPROUTE	omproute	user	None
OPORTMAP server	oportmap	daemon	None
OREXECD	rexecd	daemon	auth
ORSHD	rshd	daemon	auth
OTELNETD	telnetd	local1	auth
Policy Agent	Pagent	daemon	None
POPPER	popper	mail	None
PWCHANGE command	pwchange	daemon	None
PWTOKEY command	pwtokey	daemon	None
rpcbind	rpcbind	daemon	None
Simple Network Time Protocol daemon	sntpd	daemon	None
SNMP agent (OSNMPD)	snmpagent	daemon	None
syslogd	syslogd	daemon	None
TCP/IP subagent	M2SubA	daemon	None
TIMED daemon	timed	user	None
TN3270E Telnet subagent	TNSubA	daemon	None
Traffic Regulation Management Daemon (TRMD)	TRMD	daemon (used for IDS logging)	local4 (used for IPSEC logging and defensive filter logging) local5 (used for zERT policy-based enforcement logging)
Trap Forwarder daemon	trapfwd	daemon	None

Table 4. syslogd facilities (continued)			
Application	**syslogd record identifications**	**Primary syslog facility**	**Other syslog facility**
z/OS Load Balancing Advisor	lbadv	daemon	None
z/OS Load Balancing Agent	lbagent	daemon	None

Accounting - SMF records

Installations use Systems Management Facilities (SMF) records for the following reasons:

Performance management

Performance management includes the tasks that are related to verifying that defined service levels are met, and if not, identifying possible causes.

Aggregated information about delivered service, structured by organizational units (for which service levels have been defined) is needed to perform these tasks. These reports are typically time series with varying levels of time intervals, ranging from weeks through days to a time interval that matches the SMF interval. Some examples of potential reports related to performance management are:

- TCP connection elapsed time per server port number per time of day (potentially broken down on source IP address, or netmask)
- Number of TCP connections per server port number per time of day (potentially broken down on source IP address, or netmask)
- Number of inbound/outbound bytes transferred in TCP connections per time of day (potentially broken down in various ways: per destination or source port, per destination IP address, netmask, or in total, etc.)
- TCP retransmission activity per time of day (potentially broken down per destination IP address, or netmask)
- Number of outbound TCP connections per time of day (potentially broken down per destination IP address, or netmask)
- Number of inbound/outbound UDP datagrams per time of day (potentially broken down on server port number)
- Number of discards, error packets, and unknown protocol packets inbound and outbound per time of day (potentially broken down per interface)

Capacity planning

Capacity planning includes tasks that are related to forecasting capacity in terms of central processing power, memory, channel-based I/O subsystem, network attachments, and network bandwidth. Such planning tasks are based on analyzing trends for use of capacity during a preceding period (typically one to two years), and applying forecasting metrics, along with knowledge about planned launches of new applications or use of existing applications, to this trend in order to predict capacity needs during the next one to two year period. Some examples of potential reports related to capacity planning are:

- Total number of TCP connections per reserved server port number per day including analysis of average and variations around average during daily peak periods
- Total number of UDP inbound/outbound UDP datagrams per reserved server port number per day including average and variations around average during daily peak periods
- Number of bytes and/or packets transferred inbound and outbound per interface (LINK) per time of day (potentially broken down into unicasts, broadcasts, and multicasts)
- Size of queue length per interface per time of day

Auditing

Auditing involves tasks that are related to identifying and proving that individual events have taken place. Some examples of potential reports related to auditing are:

- Detailed information about specific TCP connections or UDP sockets, IP addresses, server/client identification, duration, number of bytes, and so on
- Details about activity that involves a specific client or server
- Details about a given application session based on server-specific SMF recording, such as individual Telnet sessions or FTP sessions
- Details about changes to the TCP/IP stack profile and the user that requested the change
- Details about changes to the status of dynamic virtual IP addresses (DVIPAs) and sysplex distributor targets

Accounting

Accounting involves tasks that are related to calculating how much each individual user or organizational unit should be charged for use of the shared central IS resources. Input to such calculations vary, but is often based on CPU cycle use, data quantities, bandwidth usage, and memory use. For TCP/IP additional metrics may be defined, such as type of service used (FTP, LPD, web server, and so on), and TCP connection-related information (number of connections, duration, byte transfer counts, and so on). Some examples of potential reports related to accounting are:

- Aggregated number of connections to a given server from a given source in terms of a specific client IP address, or netmask
- Accumulated connect time to a given server from a given source in terms of a specific client IP address, or netmask.
- Number of bytes transferred to or from a given source in terms of a specific client IP address, or netmask.
- Amount of data protected by specific manual or dynamic tunnels.
- Application-level accounting information specific to each individual server, for example:
  - For Communications Server SMTP (CSSMTP): Information about mail message processing
  - For FTP: Number of transfer operations and bytes retrieved or stored per user ID
  - For IKED: Information about IKE tunnels
  - For TN3270: Number of sessions and session type (TN3270/TN3270E/LINEMODE)

In general, SMF records are created for deferred processing and analysis and are not used for real-time monitoring purposes. In a TCP/IP environment, real-time monitoring is implemented with the SNMP protocol and is based on internal variables that SNMP subagents maintain. However, on z/OS, much of the information that is written in SMF records is also useful from a real-time monitoring perspective. For information about using z/OS Communications Server TCP/IP network management interfaces (NMIs) to obtain SMF records in real time, see z/OS Communications Server: IP Programmer's Guide and Reference.

As can be seen, all disciplines require detailed data as input. Depending on the discipline, certain levels of aggregation is performed on the raw detailed data in order to perform the tasks of that discipline. The objective of the TCP/IP product is to define and generate the lowest level of detail that is needed by all disciplines. How to aggregate and the actual aggregation is performed by other products, such as Performance Reporter for z/OS (PR), MVS Information Control System (MICS), or SAS-based tools or, in many cases, customer-written programs.

TCP/IP– produced SMF records should not be viewed isolated. Other components in MVS produce SMF records for the same purposes as those produced by TCP/IP. An installation is likely to combine information from a series of subsystems in performing detailed performance, or capacity planning. SMF records with information about use of CPU resources and memory resources per address space is, for example, produced by other components in MVS, and TCP/IP produced SMF records should not duplicate that information.

The events that trigger SMF records to be written and the information included in the SMF records must accommodate the intended purposes. There can be multiple purposes for given SMF records.

SMF records can be cut at multiple levels in the TCP/IP protocol stack, and the type of information that can be included depends on where the SMF record is created:

- At the IP and interface layer, there is information about ICMP activity, IP packet fragmentation and reassembly activity, IP checksum errors, IGMP activity, and ARP activity. At this layer, it is difficult to relate the information to specific users (remote clients, local socket address spaces, and so on), so from an accounting point of view, this information is not very interesting. Because you can aggregate network-layer activity to physical interfaces, the information at the IP and interface layer is an important aspect of both performance and capacity management.
- At the transport protocol layer, there is information about IP addresses, port numbers, and host names. For TCP-related workload, there is information about connections and information that is related to TCP connections, such as byte counts, connection times, reliability metrics, and performance metrics. For UDP-related workload, each UDP datagram is a separate entity; the only way to aggregate information for UDP is on a UDP socket level, where SMF records could be created every time a UDP socket is closed.
- At the application layer, there are more details about what goes on, but every application is different and requires separate SMF record definitions and ability to write the SMF records to implement application-layer SMF recording. Application-layer SMF recording is implemented by some applications, such as the TN3270E Telnet server (Telnet), the FTP server, the IKE daemon, and the CSSMTP daemon.

For more information about the SMF records provided by z/OS Communications Server functions, see z/OS Communications Server: IP Programmer's Guide and Reference.

Security considerations

Multilevel security is an enhanced security environment that can be configured on a z/OS system. In this environment, the security server and trusted resource managers enforce mandatory access control policies in addition to the usual discretionary access control policies. To participate in a multilevel secure environment, the user IDs associated with z/OS CS tasks and the resource profiles in the SERVAUTH class need to have security labels defined. For more information on the multilevel secure environment and configuring z/OS CS in that environment, see Chapter 4, “Preparing for IP networking in a multilevel secure environment,” on page 191.

z/OS Communications Server relies on a System Authorization Facility (SAF) to protect several resources:

- Started tasks require access to a STARTED resource. This is documented in the server information in the z/OS Communications Server: IP Configuration Reference. Also, see SEZAINST(EZARACF) for SAF authorizations required for the TCP/IP stack and servers started tasks.
- Restricting access to a network, subnetwork or particular IP address in the network is provided by resources in the SERVAUTH class. Using NETACCESS statements, z/OS CS can map networks, subnetworks and IP addresses to SAF resource names. Users that are not permitted access to a particular SAF resource are not allowed to communicate with the corresponding network, subnetwork, or IP address. See the NETACCESS statement in z/OS Communications Server: IP Configuration Reference or “Setting up the System Authorization Facility server access authorization class (optional)” on page 264 for more information.

  You can also use SAF resources in the SERVAUTH class to control which users can create dynamic VIPAs that are specified by VIPARANGE statements, and to restrict access to a specific dynamic VIPA or a specific range of dynamic VIPAs in a VIPARANGE statement. For more information about the VIPARANGE statement, see z/OS Communications Server: IP Configuration Reference. For information about VIPARANGE access control, see “Defining a security profile for SIOCSVIPA, SIOCSVIPA6, and MODDVIPA” on page 377 and “Defining a security profile for binding to DVIPAs in the VIPARANGE statement” on page 401.

  Restricting the ability of the users to run applications that access specific TCP and UDP ports is also provided by resources in the SERVAUTH class. z/OS Communications Server provides a one-to-one mapping between port numbers and SAF resource names. See the PORTACCESS statement in the z/OS Communications Server: IP Configuration Reference or “Setting up the System Authorization Facility server access authorization class (optional)” on page 264 for more information.

Similar to the function provided by the PORTACCESS statement, z/OS Communications Server ensures that a user attempting to connect to a TN3270E Telnet server secure port is allowed access to the port. This support is used in conjunction with Telnet client authentication support. See the ClientAuthType parameter of TTLSEnvironmentAdvancedParms statement in the z/OS Communications Server: IP Configuration Reference or "Setting up the System Authorization Facility server access authorization class (optional)" on page 264 for more information.

Restricting access to the TCPIP stack is also controlled under z/OS CS by defining a resource in the SERVAUTH class. See "Setting up the System Authorization Facility server access authorization class (optional)" on page 264 for more information.

- Restricting access to operator commands is provided through the OPERCMDS resource. z/OS Communications Server verifies that users have access to specific OPERCMDS resources before executing the operator command. See the operator commands information in z/OS Communications Server: IP System Administrator's Commands or "Setting up the System Authorization Facility server access authorization class (optional)" on page 264 for more information about limiting access to z/OS Communications Server commands.
- Restricting access to the TSO and UNIX shell Netstat command is provided by SERVAUTH resources. z/OS Communications Server verifies that users have access to specific SERVAUTH resources before executing the Netstat command. See the Netstat command information in the z/OS Communications Server: IP System Administrator's Commands for more information about limiting access to Netstat command. The security product resource names in the SERVAUTH class do not apply to DISPLAY TCPIP,,NETSTAT command. If you want to restrict access to DISPLAY TCPIP,,NETSTAT command, you can do so using standard operator command restriction facility, OPERCMDS class profiles. See z/OS MVS Planning: Operations for more information.

For more information about protecting these resources and other TCP/IP resources from unauthorized access, see "TCP/IP resource protection" on page 124.

Nonreusable ASIDs

The following Communications Server address spaces provide PC-entered services that must be accessible to all address spaces, so a system LX is obtained.

- Resolver
- TCP/IP stack
- TN3270 Telnet server
- VMCF and TNF subsystems

  The following applications use these subsystems:

  - SNMP Query Engine application
  - Pascal Socket Interface
  - LPD servers
  - TSO TELNET, HOMETEST, TESTSITE, RSH, REXEC, and LPR commands

Unless you specify REUSASID=YES on the START command, the Address Space Identifiers (ASIDs) associated with these address spaces will be nonreusable when these address spaces are stopped or restarted. If these address spaces are stopped enough times and you do not specify REUSASID=YES when you start the address space, all available ASIDs might be exhausted, which prevents a new address space from being created on the system. In this case, an IPL is required. Specifying REUSASID=YES when starting these address spaces ensures that the ASIDs associated with them can be reused and can help avoid an IPL. For more information about tuning parameters for the maximum number of ASIDs on a system, see the MAXUSER parameter in z/OS MVS Initialization and Tuning Reference.

Restriction: Do not specify REUSASID=YES when you are starting the VMCF and TNF subsystems or any applications that use these subsystems.

TSO command authorization

You must list the FTP, LPR, MODDVIPA, PING, and TRACERTE commands in the AUTHCMD NAMES section of your IKJTSO*xx* member of SYS1.PARMLIB.

If a command such as FTP, PING, or TRACERTE is invoked using a method other than a TSO command, you might have to do additional authorization customization, such as adding the program names to the AUTHPGM NAMES section of your IKJTSO*xx* SYS1.PARMLIB member.

UNIX System Services security considerations

This information describes some security considerations that have a product-wide effect. Descriptions of security considerations that affect specific servers or components are described with the information for each server and component.

Requirement for an OMVS segment

Many TCP/IP Services components in z/OS Communications Server now exploit z/OS UNIX services in both the native MVS environment and in the z/OS UNIX environment. For example, all TCP/IP socket APIs and TCP/IP applications (whether they are provided by z/OS Communications Server, z/OS, other IBM and non-IBM products, or written by users) now make use of z/OS UNIX services.

Use of z/OS UNIX services requires a z/OS UNIX security context, referred to as an *OMVS segment*, for the user ID associated with any unit of work requesting these services. In other words, most user IDs requiring access to TCP/IP functions now require an OMVS segment to be defined in Resource Access Control Facility (RACF).

Note: The tasks, examples, and references in this information assume that you are using the z/OS Security Server (RACF). If you are using a security product from another vendor, read the documentation for that product for instructions on task performance.

To satisfy the requirement for an OMVS segment in RACF, take one of the following actions:

- Identify all the users in your environment that use TCP/IP services, and define OMVS RACF segments with a unique UID value for the associated user IDs.
- Use the automatic assignment of unique UNIX identities support to automatically generate OMVS segments for user IDs.

For more information about automatic assignment of unique UNIX identities support, see z/OS UNIX System Services Planning. For steps to use automatic assignment of unique UNIX identities support, see z/OS Security Server RACF Security Administrator's Guide.

Authorization of TCP/IP started task user ID

The TCP/IP address space operates as a transport provider for the INET physical file system. For this to occur, the TCP/IP system address space must connect to z/OS UNIX and become a z/OS UNIX process. Therefore, the started task UID that is assigned to the TCP/IP system address space must have a valid OMVS segment.

As a transport provider, the TCP/IP address space requires superuser privileges in z/OS UNIX. Define the TCP/IP system address space started task UID as UID=0, or define the TCP/IP system address space as a trusted environment in the RACF started class profile for the TCP/IP system address space. Use the following command to assign an OMVS segment to the TCP/IP started task user ID specified as UID=0:

```
ALU tcpip_userid OMVS(UID(0) HOME(/) PGM(/bin/sh))
```

Other user IDs requiring z/OS UNIX superuser authority

When a started procedure is used to start the following servers, daemons, and agents, the user must be a superuser [UID(0)] or permitted to BPX.SUPERUSER class profile.

- Network Print Facility (NPF) queue manager
- SNMP agent (OSNMPD)

The File Transfer Protocol (FTP) daemon requires UID(0). For more information, see "Security for the FTP server" on page 666.

The following daemons are managed by the inetd server, and the user specified in file /etc/inetd.conf must be defined to RACF with UID(0). For details on inetd, see z/OS UNIX System Services Planning. For details on individual daemons, see z/OS Communications Server: IP Configuration Reference.

- z/OS UNIX remote execution daemon (REXECD)
- z/OS UNIX remote shell daemon (RSHD)
- z/OS UNIX Telnet daemon

BPX.DAEMON FACILITY class profile

Certain z/OS Communications Server TCP/IP Services servers need to change the security environment of the process in which they currently execute. For example, the FTPD daemon creates a new z/OS UNIX process for every FTP client connecting to it. After the new process is created, the daemon changes the security environment of the process so that it is associated with the security context of the logged-in user. The RACF FACILITY class resource BPX.DAEMON is used for this purpose. Table 5 on page 37 contains information about using the BPX.DAEMON resource.

Table 5. BPX.DAEMON

Task	Details
Decide if you want to activate the BPX.DAEMON level of security by reviewing the information about BPX.DAEMON authority in z/OS UNIX System Services Planning to determine whether this level of security is appropriate for your installation.	This is not required. It is recommended, however, because it provides additional security in the z/OS UNIX environment. The following TCP/IP Services servers and daemons in z/OS Communications Server change the security environment of their processes: • FTPD • Network security services (NSS) server • Policy Agent • z/OS UNIX REXECD • z/OS UNIX RSHD • z/OS UNIX TELNETD
Plan the time at which you define BPX.DAEMON carefully.	As soon as you define the BPX.DAEMON resource, MVS will not let programs change the security environment unless the programs are retrieved from a program-controlled library and unless the UID under which the program executes has access to BPX.DAEMON.
If you decide not to define the BPX.DAEMON FACILITY class profile, assign UID(0) for the UIDs associated with these servers and daemons.	This is sufficient for processing. It is described in "Other user IDs requiring z/OS UNIX superuser authority" on page 36.

Table 5. BPX.DAEMON (continued)	
Task	**Details**
If you do decide to define the BPX.DAEMON FACILITY class profile, grant READ access to this profile for the UIDs associated with the listed daemons. Also, enable BPX.DAEMON security by defining the BPX.DAEMON FACILITY class profile in RACF.	To define the BPX.DAEMON FACILITY class profile in RACF, use the following command: `RDEFINE FACILITY BPX.DAEMON UACC(NONE)` **Note:** You must specify the name BPX.DAEMON in this command. Substitutions for the name are not allowed.

If all the required conditions are not met, your server programs will fail as soon as you define BPX.DAEMON. If the server programs fail, delete BPX.DAEMON, and the setup reverts to its previous state. Check all your definitions, and make the required corrections before trying to define BPX.DAEMON again.

If this is the first FACILITY class profile that your installation is using, activate the FACILITY class using the following commands:

```
SETROPTS CLASSACT(FACILITY) GENERIC(FACILITY) AUDIT(FACILITY)
SETROPTS RACLIST(FACILITY)
```

If you start server programs using MVS start commands or from shell scripts that execute after startup of z/OS UNIX, you must allow the UIDs access to the BPX.DAEMON FACILITY class resource. The following example shows the UID (ftpd_user_ID) with which you can start the FTPD daemon:

```
PERMIT BPX.DAEMON CLASS(FACILITY) ID(ftpd_user_ID) ACCESS(READ)
```

Authorization to change the user security environment is granted only if both of the following two conditions are true:

- The server program is executing under a UID that has READ permission to the BPX.DAEMON FACILITY class profile and a UID=0.
- All programs running in the address space have been retrieved from a controlled library.

Program control

There are additional security concerns when you are loading programs that are considered trusted into the z/OS UNIX file system. Program control facilities in RACF and z/OS UNIX provide a mechanism for ensuring that the z/OS UNIX program loading process has the same security features that APF authorization provides in the native MVS environment.

It is recommended that you enable program control in your installation. If you define the BPX.DAEMON FACILITY class profile, you must enable program control for certain z/OS Communications Server load libraries. Review the information on program control in z/OS UNIX System Services Planning to decide whether program control is appropriate for your installation.

To enable program control, follow the tasks in Table 6 on page 38.

Table 6. Program control	
Task	**Details**
Activate program control.	Use the following command: `SETROPTS WHEN(PROGRAM)`

Table 6. Program control (continued)

Task	Details

Set the universal access for public library data sets (those in LINKLST*xx*) to READ. This allows access to the controlled programs and any other program in those libraries. (MVS opens the LNKLST*xx* libraries during IPL and makes these programs public. However, users cannot make changes.)	Use the following commands to create RACF data set profiles:
	`ADDSD 'cee.version.SCEERUN' UACC(READ)`
	`ADDSD 'SYS1.LINKLIB' UACC(READ)`
	`ADDSD 'TCPIP.SEZALOAD' UACC(READ)`
	`ADDSD 'TCPIP.SEZATCP' UACC(READ)`

Ensure all load modules that are loaded by the BPX.DAEMON servers into an address space come from controlled libraries.	If the MVS contents supervisor loads a module from a noncontrolled library, the address space becomes *dirty* and loses its authorization. To prevent this from happening, define all the libraries from which load modules can be loaded as program controlled. At a minimum, this should include the C run-time library, the TCP/IP Services SEZALOAD and SEZATCP libraries, SYS1.LINKLIB, and any load libraries containing FTP security exits.
	Use the following commands:
	`RDEFINE PROGRAM * ADDMEM('SYS1.LINKLIB'/volser/NOPADCHK) UACC(READ)`
	`RALTER PROGRAM * ADDMEM('SYS1.SIEALNKE'/volser/NOPADCHK)`
	`RALTER PROGRAM * ADDMEM('cee.version.SCEERUN'/volser/NOPADCHK)`
	`RALTER PROGRAM * ADDMEM('TCPIP.SEZALOAD'/volser/NOPADCHK)`
	`RALTER PROGRAM * ADDMEM('TCPIP.SEZATCP'/volser/NOPADCHK)`
	`RALTER PROGRAM * ADDMEM('db2.DSNLOAD'/volser/NOPADCHK)`
	`RALTER PROGRAM * ADDMEM('db2.DSNEXIT'/volser/NOPADCHK)`
	`RALTER PROGRAM * ADDMEM('ftp.userexits'/volser/NOPADCHK)`
	Note: If you define the load libraries as controlled, do not specify a universal access of NONE for the PROGRAM resources. If you do so for your SYS1.LINKLIB programs, you cannot IPL your z/OS system. Be aware also that in z/OS, the *volser* specification is optional.

Activate RACF changes.	Use the following command: `SETROPTS WHEN(PROGRAM) REFRESH`

Defining TCP/IP as a UNIX System Services physical file system

The TCP/IP services stack in z/OS Communications Server must be defined as a z/OS Communications Server UNIX System Services physical file system (PFS) before it can be started. This involves updating the BPXPRM*xx* parmlib member. The following sample definition in BPXPRM*xx* defines TCP/IP as a z/OS Communications Server UNIX System Services PFS, where the network layer is IP Version 4 (IPv4) and communication at the sockets layer is through the AF_INET family:

```
FILESYSTYPE TYPE(INET)  ENTRYPOINT(EZBPFINI)

NETWORK DOMAINNAME(AF_INET)
        DOMAINNUMBER(2)
        MAXSOCKETS(60000)
        TYPE(INET)
```

This sample definition shows how to define a single TCP/IP stack as IPv4 only. To define a single TCPIP stack as both IPv4 and IPv6, add an additional NETWORK statement in the BPXPRM*xx* member. The following sample definition in BPXPRM*xx* defines TCP/IP as a z/OS Communications Server UNIX System

Services PFS, where the network layer is IP Version 6 (IPv6) and communication at the sockets layer is through the AF_INET6 family:

```
NETWORK DOMAINNAME(AF_INET6)
        DOMAINNUMBER(19)
        MAXSOCKETS(60000)
        TYPE(INET)
```

The BPXPRM*xx* member contains additional parameters that are crucial to the correct operation of TCP/IP. Carefully examine and specify these parameters:

- MAXPROCSYS — Specifies the maximum number of z/OS UNIX processes that the system allows.
- MAXPROCUSER — Specifies the maximum number of processes that a single z/OS UNIX user ID can have concurrently active, regardless of how the processes were created.

  Guideline: The TN3270E server creates five UNIX processes at initialization and one more process for each BEGINVTAM statement that is used to define a listening port. For example, if 25 ports are defined by BEGINVTAM statements, then the TN3270 server creates at least 30 UNIX processes. If the TN3270E Telnet server is not running as UID(0) and has 20 or more ports defined (BEGINVTAM statements), increase the number of UNIX processes. You can increase the number of UNIX processes by increasing the value for MAXPROCUSER in the BPXPRM*xx* parmlib member, or by setting the PROCUSERMAX value in the OMVS segment for the TN3270E user ID.
- MAXUIDS — Specifies the maximum number of z/OS UNIX user IDs that can operate concurrently.
- MAXFILEPROC — Specifies the maximum number of descriptors for files, sockets, directories, and any other file system objects that a single z/OS UNIX process can have concurrently allocated. This includes access to both z/OS UNIX files and socket descriptors. In z/OS Communications Server, most TCP/IP applications access TCP/IP sockets, either directly or indirectly, using the TCP/IP socket APIs. You should set the MAXFILEPROC value high enough to accommodate the largest number of sockets that a single TCP/IP application (or z/OS UNIX process) can allocate.

  Result: The TN3270E Telnet server running as UID(0) or running with BPX.SUPERUSER authority programmatically raises this limit beyond the MAXFILEPROC value, up to the system-defined maximum.

  Guideline: If the TN3270E Telnet server or another application cannot programmatically raise this limit because the application is not running as UID(0) or with BPX.SUPERUSER authority, you can override the MAXFILEPROC limit by using the FILEPROCMAX parameter on the RACF ALTUSER command (or using the appropriate command for your security product).
- MAXPTYS — Specifies the maximum number of pseudo-terminals for the system.
- MAXTHREADTASKS — Specifies the maximum number of MVS tasks that a single process can have concurrently active.
- MAXTHREADS — Specifies the maximum number of threads that a single process can have concurrently active.
- MAXQUEUEDSIGS — The sum of MAXQUEUEDSIGS and MAXFILEPROC multiplied by 2 is the system-wide maximum for the total number of asynchronous z/OS UNIX socket calls that can be outstanding. When you are specifying this number, consider:
  - For every TCP/IP connection that the TN3270E Telnet server has, there is an asynchronous z/OS UNIX socket call outstanding, which is true for both TN3270 and TN3270E clients.
  - Any TCP/IP application, from IBM or supplied by a vendor, that uses either the z/OS UNIX asyncio callable service or the TCP/IP-provided Sockets Extended asynchronous API can have one or more outstanding asynchronous socket calls.

The MAXSOCKETS parameter specifies the maximum number of sockets that can be obtained for a particular file system type. You must ensure that this specification is large enough to accommodate your installation workload. For example, each connection to your TN3270E Telnet server or FTP server requires one z/OS UNIX System Services socket. When the maximum number of sockets is allocated, then no more Telnet sessions, FTP sessions, or other applications that require z/OS UNIX System Services sockets can be started.

Note: If multiple NETWORK statements are defined, MAXSOCKETS can be specified for each NETWORK statement and will be enforced separately.

For details on the BPXPRM*xx* member, see:

- z/OS UNIX System Services Planning
- z/OS MVS Initialization and Tuning Reference

Performance considerations

Follow the guidelines found in the z/OS MVS Initialization and Tuning Reference. If your installation is running Workload Manager, follow the guidelines found in z/OS MVS Planning: Workload Management.

VTAM, TCP/IP, and some associated server applications must be able to obtain cycles to maintain their network presence. The following dispatching priority guidelines apply for these functions:

- In general, you should set VTAM and TCP/IP to a higher dispatching priority than that of the applications that use their services.
- For server applications such as OMPROUTE, TN3270E Telnet server, IKED, NSSD, and FTPD, set the priority value to the same value to which TCP/IP is set, or to a priority that is just below that value. If you are using WLM, assign these tasks to the SYSSTC service class. Additionally, you can make these tasks non-swappable so they are available during periods of high CPU usage. The MVS default program property table sets Telnet to be non-swappable and privileged, which automatically assigns the task to the SYSSTC service class.
- Set non-critical applications, such as Policy Agent and TRMD, to a lower priority.
- Set the SNMP agent and all the SNMP subagents to the same WLM service class so that they all have the same dispatching priority. Timeouts can occur if the SNMP subagents are set to a lower dispatching priority than the SNMP agent.

On systems with significant FTP activity, you can improve performance by placing the FTP program objects into the dynamic link pack area (LPA). Putting the FTP program objects into the dynamic LPA eliminates the need to load these program objects from DASD for each FTP session. You can place these program objects into the dynamic LPA using either of the following methods:

- Include the following statement in the PROG*xx* member of SYS1.PARMLIB and then issue a SET PROG=*xx* command:

```
LPA,ADD,MODNAME(EZAFTPLS,FTPDNS,EZAFTPLC,FTP),DSNAME(LNKLST)
```

- Issue the following SETPROG command:

```
SETPROG LPA,ADD,MODNAME(EZAFTPLS,FTPDNS,EZAFTPLC,FTP),DSNAME(LNKLST)
```

Tip: You can also place the SET PROG=*xx* command or the SETPROG command in a COMMND*xx* SYS1.PARMLIB member to have the command issued at IPL time.

Requirement: If maintenance is applied to these program objects, you must update the program objects in storage using one of the following methods:

- Issue an LLA refresh command, and then issue a SET PROG=*xx* command that points to a PROG*xx* member that contains the following command:

```
LPA,ADD,MODNAME(EZAFTPLS,FTPDNS,EZAFTPLC,FTP),DSNAME(LNKLST)
```

- Issue an LLA refresh command, and then issue the following SETPROG command:

```
SETPROG LPA,ADD,MODNAME(EZAFTPLS,FTPDNS,EZAFTPLC,FTP),DSNAME(LNKLST)
```

For more information, see z/OS MVS System Commands.

Fast path support

For applications that have strict communications path length requirements, a fast path extension is provided to further reduce z/OS UNIX-to-TCP/IP stack communications. This extension is available to applications that use the UNIX System Services socket API, the z/OS XL C/C++ Runtime Library functions, and the native MVS socket APIs (such as C/C++, EZASMI macro, EZASOKET, REXX, or CICS socket APIs) provided by the Communications Server. It is not available to applications that are using the Pascal API.

With fast path support, communications are reduced for the following socket system calls:

- send()
- recv()
- sendto()
- recvfrom()
- sendmsg()
- recvmsg()

These performance enhancements are also available for the corresponding socket system calls on the z/OS UNIX asynchronous socket interface (BPX1AIO).

Rule: Automatic fast path support is activated by default.

Automatic fast path support does not have the limitations that exist with the fast path support that requires activation (see "Activating fast path support" on page 42). Most notably, automatic fast path supports POSIX signals.

Restriction: Because automatic fast path supports POSIX signals, applications that cannot handle interruptions by signals must continue to explicitly activate fast path support to use it. For more information about explicitly activating fast path support, see "Activating fast path support" on page 42.

Activating fast path support

You can activate this feature for an entire z/OS UNIX process by using the z/OS UNIX environment variable _BPXK_INET_FASTPATH. The value of this variable determines whether a socket application is marked fast path. An XL C/C++ application can set the variable by using the setenv function, or you can export the variable to the z/OS UNIX shell environment before the socket application is started. An application that is using the z/OS UNIX API can set this variable by using the BPX1ENV service.

Rule: z/OS UNIX environment variables have a process-wide scope only; that is, they usually affect a single MVS address space only. For multiple UNIX processes within a single address space, the setting of this environment variable might vary for each process within the address space. It is not a problem if some of your applications use fast path services and others do not.

If you are using the _BPXK_INET_FASTPATH environment variable, the following restrictions apply to applications that are using the UNIX System Services socket API or the z/OS XL C/C++ Runtime Library functions:

- Applications are not compliant with the X/Open Portability Guides (XPG), and POSIX signals are not supported. Applications can be interrupted only with the SIGKILL terminating signal.
- You cannot use the interactive z/OS UNIX DBX debugger to debug an application. However, you can use the DBX debugger to develop and test an application without setting the environment variable, and then run the application in production with the environment variable set.

For environments that do not use common INET, set the value of this variable to the name that is specified on the TYPE parameter of the FILESYSTYPE statement in the BPXPRMxx parmlib member.

For common INET environments, the value that is used to set the environment variable depends on whether the application is using the TCP or UDP protocols. In a common INET environment, set the variable as follows:

- For UDP applications, set it to the name of the TCP/IP stack as specified on the NAME parameter of the SUBFILESYSTYPE statement in the BPXPRMxx parmlib member. The socket application is explicitly associated with the TCP/IP stack name specified in the environment variable. This means that the socket application can communicate with accessible partners only through the specified TCP/IP stack interfaces. For UDP, the environment variable overrides the common INET support. Take this contingency into account before you activate fast path for a UDP application.

  Rule: You must use the _BPXK_INET_FASTPATH environment variable for UDP applications that require fast path support and affinity to a specific TCP/IP stack. If a UDP application establishes affinity to a specific TCP/IP stack by using other means, such as by setting the _BPXK_SETIBMOPT_TRANSPORT environment variable or by using setibmopt() or BPX1PCT, the setting of the fast path variable is ignored.
- For TCP applications, you can set the variable to an asterisk (*) to associate the application with any TCP/IP stack in the common INET configuration. This method allows all TCP/IP stacks that support the fast path model to obtain the fast path performance benefits automatically. TCP servers are not bound to a specific TCP/IP stack, even if the server specifies a specific TCP/IP stack name on the environment variable; instead, the server can listen for inbound connections across all TCP/IP stacks. When a connection arrives from the TCP/IP stack that is named in the environment variable at the time of the accept() call, it is automatically marked as fast path. Connections that arrive from TCP/IP stacks that are not named by the current environment variable value are not marked as fast path.

  Guideline: Certain TCP/IP API functions, such as the resolver services [gethostbyname(), gethostbyaddr(), getaddrinfo(), and getnameinfo()] and the network interface identification services [if_nameindex(), if_nametoindex(), and if_indextoname()] use UDP sockets internally for their processing. If a specific TCP/IP stack name is specified on the environment variable, these hidden UDP sockets are associated with only the named TCP/IP stack, which might have undesirable effects. For example, any resolver API queries that result in communications with a domain name server occur only over the specified TCP/IP stack. As a result, it is best that TCP applications set the environment variable to the special asterisk (*) value. If the application requires affinity to a specific TCP/IP stack, it can do so by using any of the facilities that are provided by z/OS UNIX, such as setibmopt() and BPX1PCT. For more information about establishing affinity to a specific TCP/IP stack, see z/OS UNIX System Services Planning.

Applications can also enable fast path processing for a single socket by issuing the Iocc#FastPath IOCTL for the socket, or by using the w_ioctl() or the BPX1IOC APIs. This IOCTL is effective only if it is issued against a socket that is already associated with a specific TCP/IP stack. Sockets are considered associated with a specific TCP/IP stack if they meet any of the following conditions:

- The application has explicit process affinity to a specific TCP/IP stack by setting the _BPXK_SETIBMOPT_TRANSPORT environment variable, or by using setibmopt(), BPX1PCT, or other methods.
- TCP/IP stack affinity is explicitly established for this socket by using the SIOCSETRTTD IOCTL.
- A bind() is already issued for the socket with a specific IP address, and not the IPv4 INADDR_ANY address or the IPv6 unspecified address, in6addr_any.
- A TCP streams socket that is connected, which includes TCP sockets that are returned as a result of accept() or sockets for which a connect() is issued.

The Iocc#FastPath constant is defined in the BPXYIOCC. This IOCTL requires a 4-byte argument as input. Set this argument to a nonzero value to activate fast path, or a zero value to disable fast path on the specified socket.

TCP receive window

The z/OS Communications Server implementation calculates the receive window as two times the receive buffer size, minus any data in the receive buffer. The calculated receive window is then limited to the value of the TCPMAXRCVBUFRSIZE parameter from the TCPCONFIG profile statement, unless Dynamic Right Sizing (DRS) is active for the connection.

DRS is a stack performance optimization that allows the TCP receive buffer size to expand because of network conditions. DRS is enabled for connections that meet the following criteria:

- The connections have inbound streaming workload.
- The round trip time (RTT) for network latency is 2 ms or longer.
- The initial receive buffer size is at least 64 KB.

When DRS is active for a connection, the advertised window follows a different set of rules:

- The maximum receive buffer size is changed to 2 MB.
- The receive buffer size is set to the estimated congestion window of the peer.

If CSM high virtual common or fixed storage is constrained and inbound data for a connection is queued in the receive buffer faster than the application can read and process it, DRS is disabled for the connection. You can determine whether DRS is active for a connection by examining the `TcpPrf` field value for the connection on the Netstat ALL/-A report output.

If a connection fails to meet the criteria for enabling DRS and thus DRS is disabled, the TCP/IP stack periodically reevaluates the connection characteristics and attempts to reactivate DRS. The TCP/IP stack activates DRS for the connection if the connection subsequently meets the criteria for DRS processing.

Guideline: When you tune for TCP data transmission, you must consider the send buffer size and the receive window size. For the z/OS Communications Server implementation, the amount of data on the network is limited by the smaller of the send buffer size and the receive window size.

TCP send window

The z/OS Communications Server implementation sets the send window to the advertised receive window of the partner. The stack uses the send window and the congestion window to control the outbound packet flow for this connection.

Use the send buffer size to determine the amount of application send data that TCP/IP can buffer for a connection. The connection send buffer size is initialized to the value that is specified on the TCPSENDBFRSIZE parameter on the TCPCONFIG profile statement. An application can explicitly set the send buffer size for a connection to a value in the valid range for the TCPSENDBFRSIZE by using SETSOCKOPT().

ORS is a stack performance optimization that allows the TCP send buffer size for a connection to expand because of network conditions. ORS is enabled for connections that meet the following criteria:

- The round trip time (RTT) for network latency is 2 ms or longer.
- The initial receive buffer size is at least 64 KB.

When ORS is active for a connection, the send buffer size is dynamically managed to optimize throughput up to maximum of 2 MB when the following criteria are met:

- CSM high virtual common or fixed storage is not constrained.
- Data that is appended to any existing data in an application send buffer exceeds the send buffer size.
- The send window and congestion window allow immediate transmission of new data that is beyond the current send buffer size.

When ORS is not active for a connection, the stack monitors the round trip time (RTT) for this connection and activates ORS if network conditions change. You can determine whether ORS is active for a connection by examining the `TcpPrf2` field value for the connection on the Netstat ALL/-A report output.

Guideline: When you tune for TCP data transmission, you must consider the send buffer size and the receive window size. For the z/OS Communications Server implementation, the amount of data on the network is limited by the smaller value of the send buffer size and the receive window size.

Outbound serialization

In some cases, z/OS Communications Server tailors its processing for certain workloads to improve in-order packet delivery on multiprocessors. This performance optimization, which is called outbound serialization, can reduce CPU consumption and improve throughput. When outbound serialization is enabled for a TCP/IP connection, outbound packets for the connection leave z/OS Communications Server in the correct order. This optimization path reduces out-of-order packet processing on the receiver and minimizes unnecessary retransmissions of packets at the sender.

A connection is eligible for outbound serialization when it meets the following criteria:

- The connection has received four or more consecutive duplicate acknowledgements (ACKs).
- The initial send buffer size is at least 64 KB.

The stack activates outbound serialization for a connection in any of the following conditions:

- The connection has been identified as operating in bulk-data fashion, such as an FTP data connection, and has been registered to the bulk-mode ancillary input queue (AIQ) of an OSA-Express feature that is enabled for inbound workload queueing.
- The connection is eligible for outbound serialization and IPSEC is also enabled for this connection.
- The connection is eligible for outbound serialization, the round trip time for the connection equals or exceeds the specified or defaulted QUEUEDRTT parameter value on the TCPCONFIG profile statement. The stack determined that a significant number of retransmissions occurred for packets that were delivered out of order initially.

Considerations for multiple instances of TCP/IP

The z/OS Communications Server TCP/IP stack is a multiple-processor capable stack, which means that it can concurrently exploit all available processors on a system. Starting multiple stacks will not yield a significant increase in throughput.

In addition, running multiple z/OS Communications Server TCP/IP stacks requires additional system resources, such as storage, CPU cycles, and DASD. It also adds a significant level of complexity to the system administration tasks for TCP/IP.

For these reasons, it is suggested that in most cases you use the INET configuration, which supports a single TCP/IP stack. However, there are some special situations where running multiple stacks can provide a benefit. For example, you might want to run two separate stacks for intranet and Internet traffic.

Common INET PFS

If you want to run multiple z/OS Communications Server TCP/IP stacks concurrently, you must use the Common INET (CINET) configuration. In this configuration, up to a maximum of eight TCP/IP stacks can be active at any time.

When the CINET configuration is used, the CINET PFS is inserted between the LFS and the TCP/IP PFS for each stack. The CINET PFS maintains an internal copy of each TCP/IP stack's IP configuration, so that it can preroute a socket call to the correct TCP/IP stack. This allows most socket programs to run with multiple stack support with no change to the application. In addition, CINET supports IPv6, and is capable of supporting underlying TCP/IP stacks in IPv4/IPv6 dual mode or in IPv4-only mode.

You can specify your choice of INET (single stack) or CINET (multiple stack) support on the NETWORK, DOMAINNAME, FILESYSTYPE, and SUBFILESYSTYPE statements of SYS1.PARMLIB(BPXPRMxx). For more information about the BPXPRMxx statements, see “Specifying BPXPRMxx values for a CINET configuration” on page 54 and z/OS UNIX System Services Planning.

Port management overview

When there is a single transport provider, and the relationship of server to transport provider is 1:1, port management is relatively simple. Using the PORT statement, the port number can be reserved for the server in the PROFILE.TCPIP for that single transport provider.

Port management becomes more complex in a CINET environment where there are multiple transport providers (multiple instances of TCP/IP) and a potential for multiple combinations of the same server (for example, z/OS UNIX and TN3270E Telnet servers).

In a multiple transport provider environment, the following questions need to be answered for each server in an installation:

- Is the server generic so that it can communicate with multiple TCP/IPs or does the server have an affinity for one instance of the transport providers and can communicate with only one TCP/IP?
- How can ports be reserved across multiple transport providers? When is the port reservation determined by MVS rather than by the job name, procedure name, or user ID?
- How can you synchronize between BPXPARMS and PORTRANGE for ephemeral port reservation?
- How can TCP/IP distinguish between two different instances of Telnet (z/OS UNIX Telnet and TN3270E Telnet servers)?

Generic server versus server with affinity for a specific transport provider

This information describes the differences between generic servers and servers with affinities for specific transport providers.

Generic server

A generic server, a server without an affinity for a specific transport provider, provides service to any client on the network. (See Figure 3 on page 46.) FTP is an example of a generic server. The transport provider is merely a connection linking client and server. The service File Transfer is not related to the internal functioning of the transport provider, and the server can communicate concurrently over any number of transport providers.

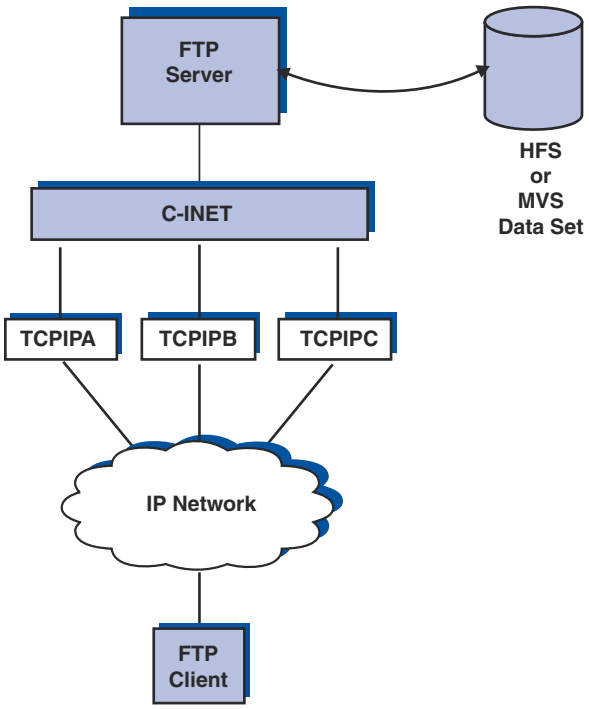


Figure 3. Generic server

Server with an affinity for a specific transport provider

When the service is related to the internal functioning of the transport provider (for example, Telnet, OMPROUTE, OSNMPD, and the Netstat command), there must be an explicit binding of the server application to the chosen transport provider. (See Figure 4 on page 47.) There must also be a way to specify the single transport to be chosen.

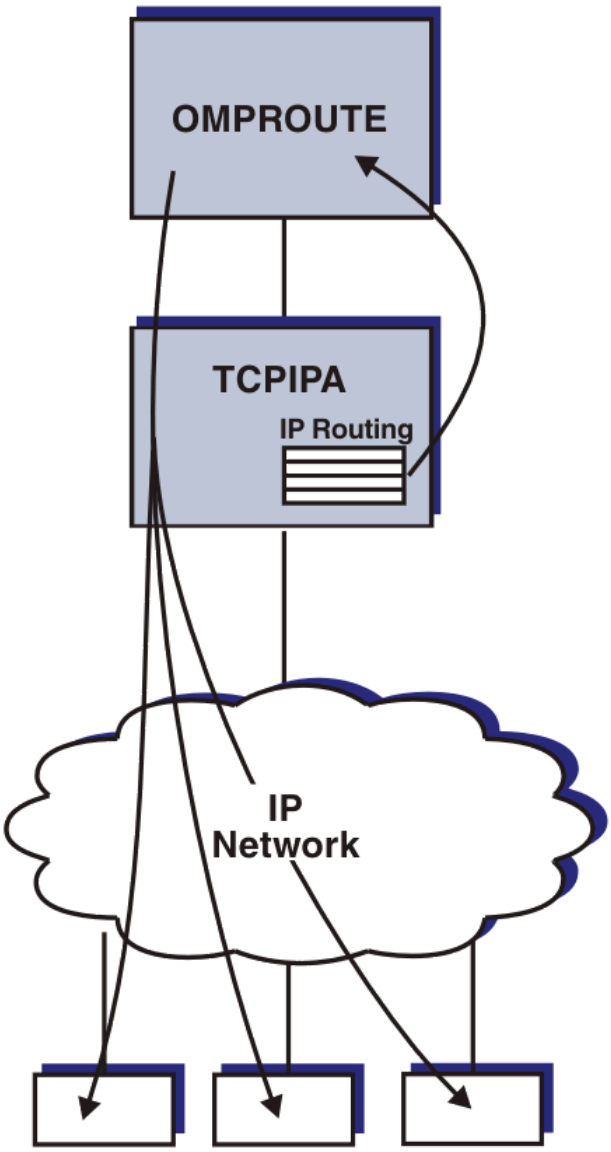


Figure 4. Server with affinity for a specific transport provider

With the exception of applications that use the socket API provided by TCP/IP, other IBM-supplied applications that use the z/OS UNIX socket API and that must bind to a specific transport provider use the z/OS UNIX socket call `setibmopt()` (see z/OS XL C/C++ Runtime Library Reference) to specify which TCP they have chosen. A C function `__iptcpn()`, described in the z/OS XL C/C++ Runtime Library Reference, enables the application to search the TCPIP.DATA file to find the name of the specific TCP/IP. (See Figure 5 on page 48.) An application that uses the z/OS Language Environment run time can also establish stack affinity by setting the environment variable _BPXK_SETIBMOPT_TRANSPORT.

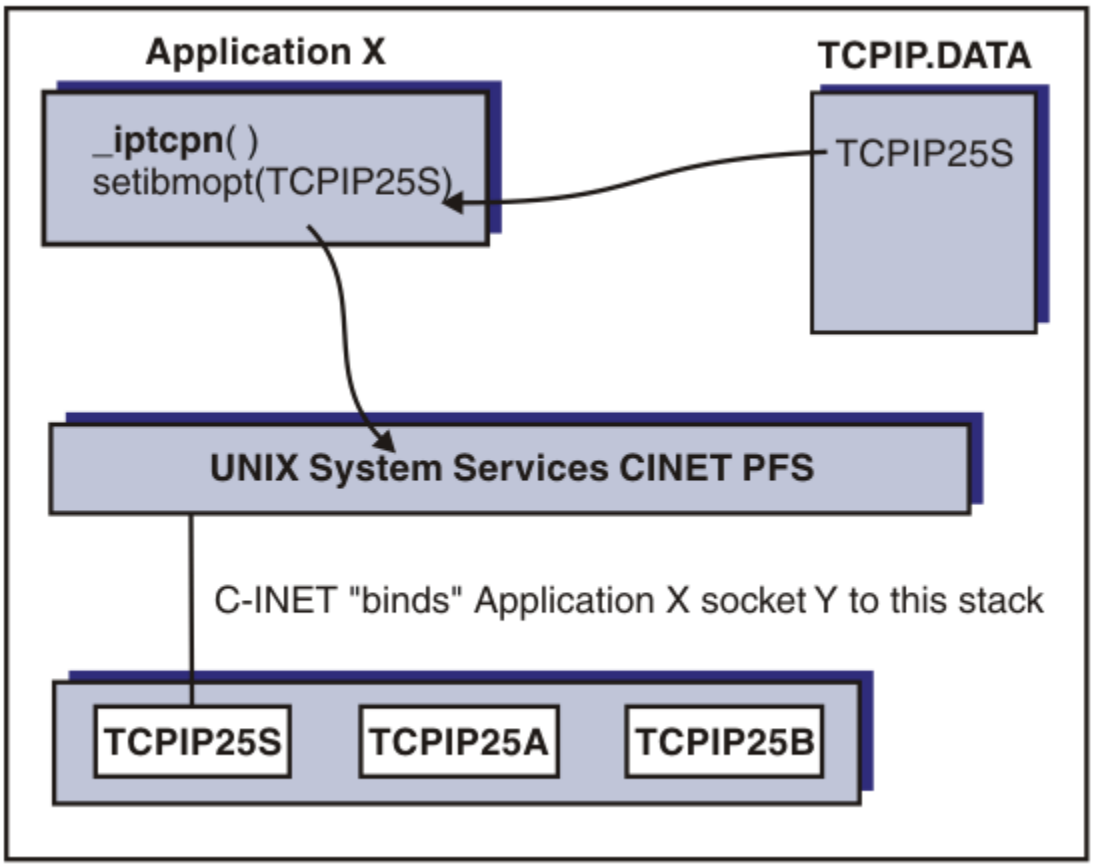


Figure 5. Example of binding an application to a specific transport provider

Generic servers in a CINET environment

In z/OS Communications Server, you can configure multiple TCP/IP stacks in a single MVS image using the CINET feature. In a CINET configuration, an application using the z/OS UNIX socket interface can get transparent access to all the TCP/IP protocol stacks configured under CINET. For example, when an application coded to z/OS UNIX sockets performs a SOCKET/BIND/LISTEN in a CINET environment, the request is propagated by CINET to all the TCP/IP stacks. This application can then service client requests that arrive into any of the configured TCP/IP stacks without having any awareness of this fact. This type of application is often referred to as a *generic server* or *daemon*.

The following servers or daemons shipped by z/OS Communications Server are generic:

- DCAS
- FTPD
- NSSD
- Policy Agent
- RPCBIND
- SNTPD
- syslogd
- TIMED
- TN3270E Telnet
- z/OS UNIX POPPER
- z/OS UNIX Portmap

- z/OS UNIX REXECD
- z/OS UNIX RSHD
- z/OS UNIX TELNETD

z/OS UNIX RSHD, REXECD and TELNETD are usually started by the INETD daemon, which is shipped as part of the z/OS UNIX. Because INETD is also a generic daemon, any server processes started by INETD inherently become generic servers as well.

If a server started by INETD (a generic server) requires affinity to a specific stack, this affinity can be accomplished by use of the _BPXK_SETIBMOPT_TRANSPORT environment variable. For more information about the _BPXK_SETIBMOPT_TRANSPORT environment variable, see z/OS UNIX System Services Planning.

The _BPXK_SETIBMOPT_TRANSPORT environment variable, when set, has an effect similar to the setibmopt() function call provided by the XL C/C++ compiler and described in the z/OS XL C/C++ Runtime Library Reference. This variable can be set in the JCL for a started procedure or batch job that executes a z/OS UNIX C/C++ program to indicate which TCP/IP stack instance the application should bind to. TCP/IP applications that require affinity to a specific TCP/IP stack, like OSNMPD and OMPROUTE, use the setibmopt() function call directly. The _BPXK_SETIBMOPT_TRANSPORT environment variable basically provides the ability to bind a generic server type of application to a specific stack.

For example, if you had two TCP/IP stacks configured under CINET, one named TCPIP and the other TCPIPOE, and you wanted to start an FTPD server instance that was associated with TCPIPOE, you could modify the FTPD procedure to set the _BPXK_SETIBMOPT_TRANSPORT environment variable as follows:

```
//FTPD   PROC MODULE='FTPD',PARMS='TRACE'
//FTPD   EXEC PGM=&MODULE,REGION=7M,TIME=NOLIMIT,
//       PARM=('POSIX(ON) ALL31(ON)',
//      'ENVAR("_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE")',
//      '/&PARMS')
//CEEDUMP  DD SYSOUT=*
//SYSFTSX DD DISP=SHR,DSN=TCPV34.STANDARD.TCPXLBIN
```

All the parameters specified prior to the slash (/) in the parameter statement are processed by the XL C/C++ run-time library. Parameters to be passed to the FTPD program must appear after the slash (/). Also note how the parameters were split over three lines in this example because they could not fit on a single line.

The following example uses JCL for the started procedure for INETD:

```
//INETD  PROC
//*************************************************************
//INETD  EXEC PGM=BPXBATCH,
//*      PARM='PGM /usr/sbin/inetd    /etc/inetd.conf'
//       PARM='PGM /usr/sbin/inetd    //''USER1.INETD.CONF'''
//*      PARM='PGM /usr/sbin/inetd    //''SYS1.TCPPARMS(INETDCNF)'''
//*
//STDERR  DD PATH='/tmp/inetd.debug.stderr',
//  PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//  PATHMODE=SIRWXU
//STDOUT  DD PATH='/tmp/inetd.debug.stdout',
//  PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//  PATHMODE=SIRWXU
//STDENV   DD  DISP=SHR,DSN=USER1.INETD.ENVIRON
//*STDENV   DD  DISP=SHR,DSN=SYS1.TCPPARMS(INETDENV)
```

The STDENV data set would contain the _BPX_SETIBMOPT_TRANSPORT variable as follows:

```
_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE
```

In these examples, INETD was also passed its configuration file as a parameter. In our examples, this file is an MVS data set rather than a z/OS UNIX file; therefore, it requires the additional double slash (//) and quotes that the example shows.

Multiple instances of INETD are not allowed, even if each instance is bound to a different TCP/IP stack. This is an INETD restriction, not a TCP/IP restriction. Therefore, if you decide to make INETD have affinity

to a specific stack, then that is the only INETD instance that you will be able to have running in that MVS image.

Notes:

1. The _BPXK_SETIBMOPT_TRANSPORT variable should be specified only for a generic server type of application.

   If specified for a non-generic server and/or non-z/OS UNIX application it will not have any effect.
2. The name specified for _BPXK_SETIBMOPT_TRANSPORT must match the job name associated with the TCP/IP stack.

   If the name specified does not match the job name of any TCP/IP stacks defined for CINET, the application will receive a z/OS UNIX return code of X'3F3' and a return value of X'005A' and may be accompanied by the following message:

   ```
   EDC8011I A name of a PFS was specified that either is
   not configured or is not a Sockets PFS.
   ```

   If the name specified does not match the job name of any currently active TCP/IP stack defined under CINET, the application will receive a z/OS UNIX return code of X'70' and a return value of X'0296' and may be accompanied by the following message:

   ```
   EDC5112I Resource temporarily unavailable.
   ```
3. For more detailed information about requesting transport affinity, see z/OS UNIX System Services Planning.

Port reservation across multiple transport providers

When there are multiple transport providers, be sure to synchronize the PORT statements in each of the PROFILE.TCPIP files to ensure that the port reservations for each stack match the port definitions for the servers that will be using that stack.

For more information about reserving ports with the PORT statement, see Chapter 5, “TCP/IP Customization,” on page 213.

Ephemeral ports

When running with multiple transport providers, just as it is necessary to synchronize PORT reservations for specific applications across all stacks, it is required to synchronize reservations for port numbers that will be dynamically assigned across all stacks. These are the ephemeral ports above 1023, which are assigned by the stack when none is specified on the application `bind()`. To reserve a group of ports in the PROFILE.TCPIP, use PORTRANGE. For more information about PORTRANGE, see Chapter 5, “TCP/IP Customization,” on page 213. Specify the same PORTRANGE for every stack. In addition, you must let the z/OS UNIX CINET know which ports are guaranteed to be available on every stack. The following example reserves ports 40000 - 41999 in the two required files:

- PROFILE.TCPIP
  - PORTRANGE 40000 2000 TCP OMVS ; Reserved for OMVS
  - PORTRANGE 40000 2000 UDP OMVS ; Reserved for OMVS
- BPXPRMxx parmlib member
  - NETWORK DOMAINNAME(AF_INET)
  - INADDRANYPORT(40000)
  - INADDRANYCOUNT(2000)

  Notes:

  1. When IPv6 is configured and there are two NETWORK statements, INADDRANYPORT and INADDRANYCOUNT need to be specified only for the NETWORK statement for AF_INET and not for

AF_INET6. If they are specified for AF_INET6, they are ignored and the values from the NETWORK statement for AF_INET are used if provided. Otherwise, the default values are used.

2. In a CINET environment, you can use IBM Health Checker for z/OS to check whether the range of ports specified by the INADDRANYPORT and INADDRANYCOUNT operands of the BPXPRM*xx* parmlib member is reserved for OMVS on the TCP/IP stack. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide.

Selecting a stack when running multiple instances of TCP/IP

Socket application programs in a multi-stack (CINET) environment must contend with the following questions:

- How the socket program selects which TCP/IP stack to use for its socket communication
- How the TCP/IP resolver code executing in the socket application address space decides which TCP/IP resolver configuration data sets to allocate

  Note: If a resolver GLOBALTCPIPDATA setup file is used, a local TCPIP.DATA cannot override any explicit statements in the global file and cannot override any resolver statements. Therefore, in a CINET environment, the TCPIPJOBNAME statement should not be specified in the GLOBALTCPIPDATA file. Also, using the GLOBALTCPIPDATA file with CINET requires that the resolver TCPIP.DATA statements are able to be used by all stacks. For example, the IP addresses specified by the NameServer statement must be accessible from all stacks. If they are not, then the GLOBALTCPIPDATA file should not be used and you should continue with multiple TCPIP.DATA data sets. For details, see Chapter 13, "The resolver," on page 729.

To answer these questions, a distinction must be made between standard servers and clients (those that come with the z/OS Communications Server product), and other socket application programs, including those you might have written yourself.

Standard servers and clients

The anchor configuration data set is the TCPIP.DATA data set. This is the base resolver configuration data set with information on host name, domain origin, and so on. It holds the TCPIPJOBNAME statement, which identifies the TCP/IP stack to use, and the DATASETPREFIX statement, which is used by the resolver code and other services when allocating configuration data sets. For more information on these data sets, see "Configuration files for TCP/IP applications" on page 25.

The key to selecting both a specific stack and resolver configuration data sets is to control which TCPIP.DATA data set a standard server or client address space allocates. Applications that use the z/OS UNIX API can use Common INET to determine which stack an application will use. But, it is important to ensure that the search order and the contents of the resolver configuration data set are understood.

Native MVS servers and clients search for TCPIP.DATA in sequences as described in "Search orders used in the native MVS environment" on page 781.

z/OS UNIX servers and clients will search for TCPIP.DATA in sequences as described in "Search orders used in the z/OS UNIX environment" on page 777.

Nonstandard servers and clients

Nonstandard servers and clients (those that do not come with the z/OS CS product) also use TCPIP.DATA to decide which resolver configuration data sets to allocate. Depending on the socket API used, they might or might not use the TCPIPJOBNAME parameter to select a stack.

If you run sockets programs from other products or vendors, you may want to know which sockets API was used to develop the program, and which techniques, if any, the program uses to specify the name of the TCP/IP system address space. As long as application programs that use a TCP/IP socket library do not specify anything specific on calls `setibmopt()`, `Initialize`, or `INITAPI`, the TCPIPJOBNAME from a TCPIP.DATA data set will be used for finding a TCP/IP system address space name.

Table 7 on page 52 depicts the differences that prevail in stack selection depending on the TCP/IP socket API under which you are running the socket program.

Table 7. How your own socket programs select a stack

C sockets	Callable and Macro	Pascal sockets	REXX sockets
SETIBMOPT or TCPIPJOBNAME from TCPIP.DATA	TCPNAME on INITAPI or TCPIPJOBNAME from TCPIP.DATA	TCPIPJOBNAME from TCPIP.DATA	Service on Initialize or TCPIPJOBNAME from TCPIP.DATA
Callable and Macro programs might have a configuration option to specify the TCP/IP system address space name, or might interrogate the available stacks with the getibmopt() call.			

A Callable or Macro program does not have to call INITAPI. If INITAPI is not called, an implicit INITAPI is performed with the value taken from TCPIPJOBNAME in a TCPIP.DATA data set. If INITAPI is called with the TCPNAME parameter specified as a space, the TCP/IP system address space name results in the TCPIPJOBNAME keyword value.

In a z/OS UNIX INET (single stack) environment, the socket application program is always associated with the single TCP/IP stack. In the z/OS UNIX Common INET (CINET) environment, your application will be associated with multiple TCP/IP stacks unless the application specifically associates with a particular stack using the z/OS UNIX socket call setibmopt(). For other ways of requesting stack affinity in a CINET environment, see z/OS UNIX System Services Planning.

TCP/IP TSO clients

TSO client functions can be directed against any of a number of TCP/IP stacks. Obviously, the client function must be able to find the TCPIP.DATA appropriate to the stack of interest at any one time. Some TSO client commands provide a parameter to specify the stack to be used. For those that do not, the following methods are available for finding the relevant TCPIP.DATA:

- Add a SYSTCPD DD statement to your TSO logon procedure. The issue with this approach is that a separate TSO logon procedure per stack is required, and users have to log off TSO and log on again using another TSO logon procedure in order to switch from one stack to another.
- Use one common TSO logon procedure without a SYSTCPD DD statement. Before a TSO user starts any TCP/IP client programs, the user has to issue a TSO ALLOC command wherein the user allocates a TCPIP.DATA data set to DD name SYSTCPD. To switch from one stack to another, the user has to deallocate the current SYSTCPD allocation (for example, TSO FREE command) and allocate another TCPIP.DATA data set.
- Combine the first and second methods. Use one logon procedure to specify a SYSTCPD DD for a default stack. To switch stacks, issue TSO ALLOC to allocate a new SYSTCPD. To switch back, issue TSO ALLOC again with the name that was on the SYSTCPD DD in the logon procedure. The disadvantage to this approach is that the name that was on the SYSTCPD DD is hidden in the logon procedure and needs to be retrieved or remembered.

The last method can be implemented by creating a small REXX program for every TCP/IP stack on your MVS system. For each stack create a REXX program with the name of the stack (for example, T18A or T18B). Whenever TSO users want to use the T18A stack, they run the T18A REXX program. Any TCP/IP functions invoked thereafter will use the T18A stack for socket communication. If users want to switch to the T18B stack, they run the T18B REXX program. See Figure 6 on page 53 for an example.

```
/* REXX  "T18B" */
/*********************************************************************/
/*                                                                   */
/* Switch TSO Address Space to use the T18B Stack.                   */
/* Subsequent NETSTAT command will be directed toward                */
/*  the T18BTCP stack.                                               */
/*                                                                   */
/*********************************************************************/
Say 'Switching to T18BTCP stack'

msgstat = msg()
z = msg("OFF")
"FREE FI(SYSTCPD)"
"ALLOC FI(SYSTCPD) DA('TCPIP.T18B.TCPPARMS(TCPDATA)') SHR"
z = msg(msgstat)

exit(0)
```

Figure 6. REXX program to switch TSO user to another TCP/IP stack

Selecting configuration data sets

The resolver code and other services that execute as part of the socket program address space to service calls such as `gethostbyname()`, `getservbyname()` and `getprotobyname()` allocate one or more resolver configuration files to service these calls. All socket programs, including standard servers and clients and homegrown socket programs, need access to resolver configuration files. For information on how the resolver configuration files are found and used, see "Configuration files for TCP/IP applications" on page 25.

Sharing resolver configuration data sets

The general recommendation is to use separate DATASETPREFIX values for each stack and create separate copies of the required configuration data sets; at the very least, create separate copies of the resolver configuration data sets. For a test and a production stack, however, you would probably use different DATASETPREFIX values. However, if the stacks are functionally identical, you may share the same DATASETPREFIX values and many of the same configuration data sets. You need separate TCPIP.DATA data sets because of the two different TCPIPJOBNAMEs. On the other hand, you may choose to share the resolver configuration data sets between the stacks by using the same DATASETPREFIX value in each TCPIP.DATA data set.

In addition to separate TCPIP.DATA data sets, separate /etc/resolv.conf files might also be necessary. If this is the case, use the environment variable RESOLVER_CONFIG to point to the appropriate resolver information.

Exercise caution if servers use DATASETPREFIX to allocate server-specific configuration data sets. Try to use explicit allocation as far as possible in your server JCL procedures. Most servers allow you to explicitly allocate their configuration data sets using DD statements.

Some servers may use DATASETPREFIX to create new data sets. Servers that do create new data sets allow you to specify an alternate data set prefix for the data sets that are created. NPF creates new sequential data sets with captured print data. NPF has a special keyword in NPF.DATA for this purpose; it is called NPFPRINTPREFIX. If this keyword is specified, NPF will use that as the high-level qualifier for newly created print data sets instead of taking the DATASETPREFIX value from TCPIP.DATA.

Specifying BPXPRMxx values for a CINET configuration

For a detailed description of parameters in SYS1.PARMLIB(BPXPRMxx), see z/OS UNIX System Services Planning and z/OS MVS Initialization and Tuning Guide.

```
/* AF_INET file system for sockets      */
/* CINET support - BPXTCINT  */

FILESYSTYPE TYPE(CINET)
            ENTRYPOINT(BPXTCINT) 1
NETWORK DOMAINNAME(AF_INET)
        DOMAINNUMBER(2)
        MAXSOCKETS(10000) 2         TYPE(CINET)
        INADDRANYPORT(40000) 3
        INADDRANYCOUNT(2000)
NETWORK DOMAINNAME(AF_INET6)4
        DOMAINNUMBER(19)
        MAXSOCKETS(10000) 2         TYPE(CINET)
SUBFILESYSTYPE NAME(TCPIP1A) 5
               TYPE(CINET)
               ENTRYPOINT(EZBPFINI) 6
               DEFAULT 7
SUBFILESYSTYPE NAME(TCPIP1B) 5
               TYPE(CINET)
               ENTRYPOINT(EZBPFINI)
```

Figure 7. SYS1.PARMLIB(BPXPRMxx) for CINET

Notes:

1. CINET and BPXTCINT specify the use of CINET.
2. The MAXSOCKETS operand specifies the maximum number of sockets that can be obtained for the given file system type. It should be large enough for the number of sockets needed for applications using z/OS Communications Server. MAXSOCKETS is enforced independently for AF_INET (IPv4 sockets) and AF_INET6 (IPv6 sockets).
3. The INADDRANYPORT and INADDRANYCOUNT operands specify the first ephemeral port number and the range of ports to be used by z/OS UNIX CINET. The port range specified should also be reserved for CINET use in the TCP/IP profile using the port reservation statements. For details, see "Port reservation across multiple transport providers" on page 50. You can use IBM Health Checker for z/OS to check whether the range of ports specified is reserved for OMVS on the TCP/IP stack. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide.
4. This additional NETWORK statement is required if you want a TCP/IP stack to also support IPv6. Omit this statement if you do not want the stack to support IPv6 (that is, the stack will support IPv4 only).
5. A transport provider stack for CINET is specified with a SUBFILESYSTYPE statement. The NAME field must match the address space name for the TCP/IP started task as well as the TCPIPJOBNAME parameter in TCPIP.DATA. In our example, the name of the first stack is TCPIP1A and the name of the second stack is TCPIP1B.
6. EZBPFINI identifies a z/OS Communications Server TCP/IP stack. For a z/OS Communications Server TCP/IP stack, this is the only valid value.
7. Keyword DEFAULT specifies which transport provider stack is to be used as the default stack for z/OS UNIX. If DEFAULT is not specified, the first active stack will be used as the default stack. The sequence of SUBFILESYSTYPE statements is arbitrary if one stack is identified with the keyword DEFAULT. TCPIP1A is the default stack in Figure 7 on page 54.

Considerations for Enterprise Extender

The Enterprise Extender (EE) network connection is a simple set of extensions to the existing open high-performance routing (HPR) technology. It performs an efficient integration of the HPR frames using UDP/IP packets. To the HPR network, the IP backbone is a logical link. To the IP network, the SNA traffic is UDP datagrams that are routed without any hardware or software changes to the IP backbone. Unlike gateways, there is no protocol transformation and unlike common tunneling mechanisms, the integration

is performed at the routing layers without the overhead of additional transport functions. The advanced technology enables efficient use of the intranet infrastructure for support of IP-based client accessing SNA-based data (for example, Telnet emulators or web browsers using services such as IBM's Host On-Demand) as well as SNA clients using any of the SNA LU types.

Enterprise Extender seamlessly routes packets through the network protocol *edges*, eliminating the need to perform costly protocol translation and the store-and-forward associated with transport-layer functions. Unlike Data Link Switching (DLSw), for example, there are no TCP retransmit buffers and timers and no congestion control logic in the router because it uses connectionless UDP and the congestion control is provided end system to end system. Because of these savings, the *edge* routers have less work to do and can perform the job they do best, which is forwarding packets instead of incurring protocol translation overhead and maintaining many TCP connections. Data center routers can handle larger networks and larger volumes of network traffic, thus providing more capacity.

Enterprise Extender supports both the IPv4 and IPv6 addressing models. For more information about EE, see z/OS Communications Server: SNA Network Implementation Guide and the EE information in *Migrating Subarea Networks to an IP Infrastructure Using Enterprise Extender* (IBM Redbooks).

You can use QDIO inbound workload queueing on an OSA-Express3 feature (or later feature) to improve throughput for both inbound Enterprise Extender (EE) packets and for inbound non-EE traffic over the same interface. For more information, see "QDIO inbound workload queueing" on page 71.

Considerations for VIPA

The Internet Protocol (IP) is a connectionless protocol. IP packets are routed from the originator through a network of routers to the destination. All physical adapter devices in such a network, including those for client and server hosts, are identified by an IP Address which is unique within the network. The important point about IP is that a failure of an intermediate router node or adapter will not prevent a packet from moving from source to destination, as long as there is an alternate path through the network.

TCP sets up a connection between two endpoints, identified by the respective IP addresses and a port number on each. Unlike failures of an adapter in an intermediate node, if one of the endpoint adapters (or the link leading to it) fails, all connections through that adapter fail and must be reestablished. If the failure is on a client workstation host, only the relatively few client connections are disrupted and usually only one person is inconvenienced. However, an adapter failure on a server means that hundreds or thousands of connections may be disrupted. On an S/390® or System z server with large capacity, the number may run to tens of thousands.

A Virtual IP Address, or VIPA in TCP/IP for z/OS , alleviates this situation. A VIPA is configured in the same way as a normal IP address for a physical adapter, except that it is not associated with any particular device. To an attached router, the TCP on z/OS simply looks like another router. When the TCP receives a packet destined for one of its VIPAs, the inbound IP function of the stack notes that the IP address of the packet is in the stack's Home list and passes the packet up the stack. Assuming the stack has multiple adapters or paths to it (including XCF from other TCP stacks in a sysplex), if a particular physical adapter fails, the attached routing network will route VIPA-targeted packets to the stack using an alternate route.

While this removes hardware and associated transmission media as a single point of failure for large numbers of connections, the connectivity of a server can still be lost through a failure of a single stack or an MVS image. The VIPA can be configured on another stack with a manual process, but this requires the presence of an operator or programmed automation.

Dynamic VIPA Takeover enables Dynamic VIPAs to be moved without human intervention or programmed automation to allow new connections to a server at the same IP address as soon as possible. This can reduce downtime significantly. With Dynamic VIPA Takeover you can configure one or more TCP/IP stacks to be backups (VIPABACKUP statement) for a particular Dynamic VIPA. If the stack or MVS image where the Dynamic VIPA is active is terminated, one of the backup stacks automatically activates that Dynamic VIPA. The existing connections will be terminated but can be quickly reestablished on the stack that is taking over.

Notes:

1. Because a VIPA is associated with a z/OS TCP/IP stack and is not associated with a specific physical network attachment, it can be moved to a stack on any image in the sysplex, or even to a z/OS TCP/IP stack not in the sysplex as long as the address fits into the installation's network configuration.
2. If using VIPA along with an intelligent bridge or switch, ensure that **Port fast mode** (Cisco) is enabled. This helps to decrease the amount of time the VIPA is unreachable in scenarios where there is dynamic movement of VIPA (dynamic or static). For more information, see your bridge or switch manual.

You may also associate a particular Dynamic VIPA address with an application using the SIOCSVIPA or SIOCSVIPA6 ioctl command or by BINDing explicitly to the Dynamic VIPA address. If the Dynamic VIPA address is within the VIPARANGE profile statement, then this Dynamic VIPA address will be created dynamically. This type of configuration enables a Dynamic VIPA to become an address of an application in a sysplex.

With sysplex distributor you can distribute connection requests destined for Dynamic VIPAs to other stacks in the sysplex. You can use the VIPADISTRIBUTE profile statement to designate up to 32 stacks and 64 ports where connections for a particular DVIPA can be distributed, including the stack where the DVIPA is defined.

The distributing stack (the stack where the VIPADISTRIBUTE statement was coded) might use either WLM or a combination of WLM and Quality of Service (QoS) performance information to determine where to forward new connection requests. If the distributing stack/MVS image fails, connections forwarded to target stacks can be preserved by having the Dynamic VIPA address backed up on another stack.

Similarly, a stack can immediately take back a Dynamic VIPA address from another stack. If the original stack used an address specified with VIPADEFINE with the keyword MOVEABLE IMMEDIATE (the default), then the Dynamic VIPA is moved as soon as the second stack requests ownership. The second stack assumes responsibility for forwarding packets for existing connections to the appropriate stack. If MOVEABLE WHENIDLE was specified, ownership does not pass until all existing connections on the current stack are closed.

For detailed information about VIPA, see Chapter 7, “Virtual IP Addressing,” on page 361.

Considerations for Fast Response Cache Accelerator

Fast Response Cache Accelerator (FRCA) is a Communications Server function that can significantly improve the performance of the z/OS HTTP Server and the WebSphere® Application Server on z/OS. Web pages are cached within the operating system kernel and requests are handled without traversing the entire kernel or entering the user space. For more information about configuring the WebSphere Application Server to use the FRCA function, see the WebSphere Application Server documentation at http://www.ibm.com/support/knowledgecenter/SSEQTP/mapfiles/product_welcome_was.html.

FRCA also provides the ability to perform content-based quality-of-service (QoS) classification by selecting an appropriate QoS policy for each individual URI in the HTTP request. For more information on specifying individual URIs, see the Policy Agent and policy applications topic in z/OS Communications Server: IP Configuration Reference.

Use the Netstat CACHINFO/-C commands to display information about FRCA statistics. Statistics are displayed for each listening socket configured for FRCA support. For more information on the Netstat CACHINFO/-C reports, see z/OS Communications Server: IP System Administrator's Commands.

Considerations for extended address volumes

As of Version 1 Release 10, z/OS supports extended address volumes (EAVs). EAVs are volumes greater than 65,520 cylinders. The part of an EAV beyond the first 65,536 cylinders is called the extended address space (EAS). Not all data sets on an EAV can be in the EAS. A data set that could be in the EAS, whether it actually is in the EAS or not, is said to be an EAS-eligible data set.

Rule: Do not allocate data sets that are used by Communications Server as EAS-eligible data sets, except as follows:

- The FTP client and server allow transfer to or from the following types of EAS-eligible data sets:
  - Physical sequential basic format data sets
  - Physical sequential large format data sets
  - Physical sequential extended format data sets
  - Partitioned data sets (PDSs)
  - Libraries
- The following FTP configuration files can be EAS-eligible data sets:
  - ANONYMOUSLOGINMSG
  - ANONYMOUSMVSINFO
  - BANNER
  - FTP.DATA

    Restriction: The TSO HOMETEST command cannot process FTP.DATA when it is an EAS-eligible data set.
  - LOGINMSG
  - MVSINFO
  - NETRC
  - SOCKSCONFIGFILE

For more information about using extended address volumes, see z/OS DFSMS Using the New Functions.

Considerations for networking hardware attachment

This information provides general networking hardware attachment information and considerations associated with the IBM Z® platform. Most of the information included here is associated with the IBM Open Systems Adapter (OSA) and the QDIO system architecture.

OSA-Express feature in QDIO mode

When an OSA is configured with a DEVICE and LINK definition, the z/OS TCP/IP stack has only one connection to the OSA-Express feature. You define this connection with a combination of the DEVICE, LINK, and HOME statements for IPv4, an INTERFACE statement for IPv6, or both. This single connection uses one channel unit address for communication, which is assigned by VTAM from the DATAPATH parameter of the TRLE definition that represents this OSA-Express feature. Both IPv4 and IPv6 traffic share this one connection and channel unit address. For an overview of the OSA-Express feature and QDIO mode, and how to configure VTAM to use the OSA-Express feature in this mode, see z/OS Communications Server: SNA Network Implementation Guide.

When you configure an OSA-Express feature for IPv4 using the INTERFACE statement and also configure the same OSA-Express feature for IPv6, then the z/OS TCP/IP stack has two connections to the OSA-Express feature and requires two DATAPATH devices.

Steps for converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the IPv4 IPAQENET INTERFACE statement

Use the INTERFACE statement to configure IPv4 definitions for OSA-Express QDIO, rather than using the DEVICE, LINK, and HOME statements. The INTERFACE statement improves stack configuration for IPAQENET interfaces, and some functions like multiple VLAN support require that the QDIO interface is defined with the INTERFACE statement.

Procedure

Perform the following steps to convert your TCP/IP profile so that it uses the INTERFACE statement to configure IPv4 definitions for OSA-Express QDIO.

1. Convert the IPv4 IPAQENET DEVICE, LINK, and HOME statements to an IPv4 IPAQENET INTERFACE statement.

   The values used in the following sub-steps are based on the sample profile that appears after the sub-steps.

   a. Copy the LINK name (QDIO4101L in the example) and specify this name as the interface name on the INTERFACE statement.

      Tip: When you use the original LINK name as the new INTERFACE name, you do not have to make changes to the static route definitions, OMPROUTE definitions, PKTTRACE statement, and PRIMARYINTERFACE statement.

   b. Copy the IPAQENET parameter from the LINK statement and specify this parameter after the DEFINE parameter of the INTERFACE statement.

   c. Copy the remaining LINK parameters and values (INBPERF DYNAMIC in the example) to the INTERFACE statement.

   d. Copy the DEVICE name (QDIO4101 in the example) and specify this name on the PORTNAME parameter of the INTERFACE statement. This PORTNAME value must match the corresponding PORTNAME value in the SNA TRLE definition.

   e. Copy the remaining DEVICE parameters (PRIROUTER in the example) to the INTERFACE statement.

   f. Copy the HOME list entry IP address (172.16.1.1 in the example) and specify this address on the IPADDR parameter of the INTERFACE statement.

   g. Append a value (/24 in the example) to the end of the IP address to define the subnet mask.

   h. Remove the HOME list entry.

   Example of the profile before conversion:

```
DEVICE QDIO4101  MPCIPA   PRIROUTER
LINK   QDIO4101L IPAQENET      QDIO4101
  INBPERF DYNAMIC
;
HOME
  172.16.1.1     QDIO4101L
```

   Example of the profile after conversion:

```
INTERFACE QDIO4101L
  DEFINE IPAQENET
  INBPERF DYNAMIC
  IPADDR 172.16.1.1/24
  PORTNAME QDIO4101
  PRIROUTER
```

   Tip: Optionally, you can dump the TCP/IP address space and use the CONVERT parameter on the TCPIPCS PROFILE subcommand to display the configuration information at the time of the dump. The resulting output reflects your IPAQENET DEVICE, LINK, and HOME definitions in INTERFACE statement format, which might be helpful in converting your profile to use INTERFACE statements. You should review the output before you implement any changes. For more information about using the CONVERT parameter on the TCPIPCS PROFILE subcommand, see z/OS Communications Server: IP Diagnosis Guide.

   For more information about the IPv4 IPAQENET INTERFACE statement, see z/OS Communications Server: IP Configuration Reference.

2. Remove any BSDROUTINGPARMS entries for the interface.
3. If you have configured the SOURCEVIPA parameter on the IPCONFIG statement, perform the following steps:

   a) Find the IPAQENET LINK in the original HOME list and search backwards to locate the static VIPA (if any) that is located closest to this link in the HOME list.

   b) If you find a static VIPA, add the SOURCEVIPAINTERFACE parameter to the IPv4 INTERFACE statement. Use the static VIPA link name as the SOURCEVIPAINTERFACE value.

4. If you are using the START statement to start the IPv4 device, change the START statement to specify the name of the IPv4 INTERFACE.
5. If you also have an IPAQENET6 definition for the OSA, perform the following steps:
   a) If you configure the same virtual MAC address (VMAC) on both the IPAQENET LINK statement and IPAQENET6 INTERFACE statement, either change one of the VMAC addresses so that they are unique or let OSA generate the VMAC addresses.
   b) Ensure that the corresponding TRLE definition has at least two DATAPATH devices available so that one device is available for the IPv4 interface and one device is available for the IPv6 interface.

Results

For information about the TCP/IP profile (PROFILE.TCPIP) and configuration statements, see z/OS Communications Server: IP Configuration Reference.

The following examples show some additional changes that you might need to make to the definitions for OSA-Express QDIO.

Example of the profile before conversion:

```
IPCONFIG
  SOURCEVIPA
;
DEVICE VIPA4811  VIRTUAL  0
LINK   VIPA4811L VIRTUAL  0    VIPA4811
;
DEVICE QDIO4101  MPCIPA   PRIROUTER
LINK   QDIO4101L IPAQENET      QDIO4101
  INBPERF DYNAMIC
;
HOME
  10.81.1.1      VIPA4811L
  172.16.1.1     QDIO4101L
;
PRIMARYINTERFACE QDIO4101L
;
BSDROUTINGPARMS TRUE
  VIPA4811L 1492 0 255.255.255.0 0
  QDIO4101L 1492 0 255.255.255.0 0
ENDBSDROUTINGPARMS
;
BEGINROUTES
  ROUTE 172.16.0.0/24  =  QDIO4101L  MTU 1492
ENDROUTES
;
START QDIO4101
```

Example of the profile after conversion:

```
IPCONFIG
  SOURCEVIPA
;
DEVICE VIPA4811  VIRTUAL  0
LINK   VIPA4811L VIRTUAL  0    VIPA4811
;
; Converted INTERFACE statement
;
;   - QDIO4101L is from the LINK statement
;   - DEFINE IPAQENET is from the LINK statement
;   - INBPERF DYNAMIC is from the LINK statement
;   - IPADDR 172.16.1.1 is from the HOME list entry, /24 is from the
;     BSDROUTINGPARMS entry subnet mask
;   - PORTNAME QDIO4101 is from the DEVICE statement
;   - PRIROUTER is from the DEVICE statement
;   - SOURCEVIPAINTERFACE VIPA4811L is from the order of the HOME list
;     entries
;
INTERFACE QDIO4101L
  DEFINE IPAQENET
  INBPERF DYNAMIC
  IPADDR 172.16.1.1/24
  PORTNAME QDIO4101
  PRIROUTER
  SOURCEVIPAINTERFACE VIPA4811L
```

```
;
; QDIO4101L is removed from the HOME list
;
HOME
  10.81.1.1      VIPA4811L
;
PRIMARYINTERFACE QDIO4101L
;
; QDIO4101L is removed from BSDROUTINGPARMS
;
BSDROUTINGPARMS TRUE
  VIPA4811L 1492 0 255.255.255.0 0
ENDBSDROUTINGPARMS
;
BEGINROUTES
  ROUTE 172.16.0.0/24  =  QDIO4101L  MTU 1492
ENDROUTES
;
; START statement uses the interface name
;
START QDIO4101L
```

Virtual LAN

A local area network (LAN) is a broadcast domain. Nodes on a LAN can communicate with each other without a router, and nodes on different LANs need a router to communicate. A virtual LAN (VLAN) is a configured logical grouping of nodes using switches. Nodes on a VLAN can communicate with each other as if they were on the same LAN, and nodes on different VLANs need a router to communicate.

OSA VLAN

The IBM Open Systems Adapter provides support for IEEE standards 802.1p/q, which describes priority tagging and VLAN identifier tagging. Deploying VLAN IDs allows a physical LAN to be partitioned or subdivided into discrete virtual LANs. This support is provided by the z/OS TCP/IP stack and the OSA-Express feature in QDIO mode.

When you use VLAN IDs, the z/OS TCP/IP stack can have multiple connections to the same OSA-Express feature. One connection is allowed for each unique combination of VLAN ID and IP version (IPv4 or IPv6). Each connection is defined by an INTERFACE statement and uses one channel unit address for communication, which is assigned by VTAM from the DATAPATH parameter of the TRLE definition. To configure one or more VLANs for a single OSA-Express feature, take the following steps:

- Configure each IPv4 interface for this OSA-Express feature in the TCP/IP profile by using the INTERFACE statement for IPAQENET, rather than the DEVICE, LINK, and HOME statements. Configure each IPv6 interface for this OSA-Express feature in the TCP/IP profile by using the INTERFACE statement for IPAQENET6.
- Configure a VLANID value on each IPv4 INTERFACE statement and each IPv6 INTERFACE statement on this stack for this OSA-Express feature. Within each IP version, these VLANID values must be unique.
- Configure the VMAC parameter on each of the INTERFACE statements with the default ROUTEALL attribute. You can specify the VMAC address or OSA can generate it. If you specify a VMAC address, it must be unique for each INTERFACE statement.
- Configure a unique subnet for each IPv4 interface for this OSA-Express feature, by using the subnet mask specification on the IPADDR parameter on the INTERFACE statement.
- If you are using OMPROUTE and OMPROUTE is not configured to ignore this interface, ensure that the subnet mask value that you configure on the INTERFACE statement in the TCP/IP profile matches the subnet mask that is used by OMPROUTE for this interface. The subnet mask that OMPROUTE uses is the subnet mask value that is defined on the corresponding OMPROUTE statement (OSPF_INTERFACE, RIP_INTERFACE, or INTERFACE) for this interface. If no OMPROUTE statement is specified for this interface, the subnet mask that OMPROUTE uses is the class mask for the interface IP address.
- Ensure that a DATAPATH device is available in the TRLE definition for each IPv4 interface and for each IPv6 interface for this OSA-Express feature.

- Ensure that you understand the fixed storage requirements for this configuration; each interface requires its own DATAPATH device, and each DATAPATH device requires fixed storage for read processing. For more information, see "Fixed storage requirements for OSA-Express QDIO and HiperSockets interfaces" on page 106.

Restrictions:

- The stack supports a maximum of 32 VLAN interfaces per IP version to the same OSA-Express port.
- For a particular OSA-Express feature, a device and link definition is precluded in the following scenarios:
  - An IPv4 interface is or was previously defined.
  - Multiple IPv6 definitions are or were previously defined.

OSA routing

For QDIO devices, z/OS Communications Server and the OSA-Express feature provide functions that control how incoming unicast datagrams are routed, especially when the OSA-Express feature is shared by multiple TCP/IP instances.

When TCP/IP activates a QDIO device, each TCP/IP registers each of its home IP addresses with the OSA-Express feature. (TCP/IP also dynamically registers any updates to its set of home IP addresses with the OSA-Express feature.) This enables the OSA-Express feature to route datagrams destined for a registered IP address to the correct TCP/IP instance.

However, when packets are received for IP addresses that are not registered by any TCP/IP stack, the device needs a method for determining which stack, if any, should receive the packet. Two functions are available to accomplish this:

- OSA-Express Virtual MAC (VMAC) routing
- Primary and secondary routing

OSA-Express virtual MAC routing

If multiple TCP/IP instances are sharing an OSA-Express feature, the preferred method of routing is to define or generate a virtual MAC (VMAC) for each stack and for each protocol being used (IPv4 or IPv6). For IPv4, this results in the OSA-Express feature using the VMAC address rather than the physical *burned in* MAC for all ARPs sent for that TCP/IP stack's registered IP addresses, and using the VMAC as the source MAC address for all packets sent from that stack. In this way, all routers on the same LAN as the OSA-Express feature use only the VMAC address as the destination for all packets destined for that specific TCP/IP stack. From a network routing perspective, the OSA-Express feature with this VMAC appears as a dedicated device to that TCP/IP stack.

This simplifies a shared OSA configuration significantly. The routers on the LAN always send any packets destined for a particular TCP/IP stack to the VMAC defined for that stack. The OSA-Express feature knows by VMAC address exactly which stack should receive a given packet. Even if the IP address is not registered with the OSA-Express feature, if the packet is destined for that VMAC, the router has determined which stack should be the intermediate router, and the OSA can forward the packet directly to that stack. If the stack is not an intermediate router, the capability is provided for a stack to indicate to the OSA that it wants to receive packets to registered IP addresses only.

This simplification is true for IPv6 as well. TCP/IP uses the VMAC address for all neighbor discovery address resolution flows for that stack's IP addresses, and likewise uses the VMAC as the source MAC address for all IPv6 packets sent from that stack. Again, from a network perspective, the OSA-Express feature with this VMAC appears as a dedicated device to that stack.

The VMAC address can be defined in the stack, or it can be generated by the OSA. If generated by the OSA, it is guaranteed to be unique from all other physical MAC addresses and from all other VMAC addresses generated by any OSA-Express feature.

Rule: If VMACs are defined in the stack, they should be defined as locally administered MAC addresses, and should be unique addresses for the local LAN on which they are.

Guidelines:

- If the OSA is configured for both IPv4 (using DEVICE and LINK) and IPv6 for a stack, then you can define the same VMAC for both the INTERFACE statement and the LINK statement, or you can define one VMAC on the LINK statement for IPv4 usage and a different VMAC on the INTERFACE statement for IPv6 usage. If the OSA is configured for both IPv4 (using the INTERFACE statement) and IPv6 for a stack, then you must define one VMAC on the INTERFACE statement for IPv4 usage, and a different VMAC on the INTERFACE statement for IPv6 usage.
- A VLAN ID can be associated with an OSA-Express link or interface that is defined with a VMAC. For more information about VMAC routing, see z/OS Communications Server: SNA Network Implementation Guide.

To enable virtual MAC support, you must be running at least an IBM System z9® Enterprise Class (z9® EC) or z9 Business Class (z9 BC), and an OSA-Express feature with OSA Layer 3 Virtual MAC support. OSA Layer 3 Virtual MAC support is not available for Fast Ethernet. For more information, see the 2094DEVICE Preventive Service Planning (PSP) bucket and the 2096DEVICE Preventive Service Planning (PSP) bucket.

Primary router

If only one TCP/IP instance is using the OSA-Express device, or when multiple TCP/IP instances are using the same OSA-Express device but you want all instances to share the same physical MAC address of the device, you can optionally have the OSA route unicast packets to unregistered IP addresses by using OSA's primary (PRIROUTER) and secondary (SECROUTER) router support. This function enables a single TCP/IP stack, on a per-protocol (IPv4 and IPv6) basis, to register and act as a router stack on a per-OSA basis. Secondary routers can also be configured to provide for conditions in which the primary router becomes unavailable and the secondary router takes over for the primary router. The primary router stack is the only stack to which OSA forwards packets when the destination IP address has not been previously registered with OSA. If no active TCP/IP instance using this device is defined as the primary router (PRIROUTER) or secondary router (SECROUTER), the device discards the datagram. If the PRIROUTER or SECROUTER value is not specified, the default value is NONROUTER.

Relationship of VLAN and primary router

The OSA primary router support takes into consideration and interacts with the VLAN ID support (VLAN ID registration and tagging). OSA supports a primary and secondary router on a per VLAN basis (per registered VLAN ID). Therefore, if TCP/IP is configured with a specific VLAN ID and also configured as a primary or secondary router, that stack serves as a router for just that specific VLAN. This allows each OSA (CHPID) to have a primary router per VLAN. Configuring multiple primary routers (one per VLAN) has many advantages and preserves traffic isolation for each VLAN.

This support becomes more important when a single OSA is shared by multiple stacks. In this type of configuration, when each stack was configured with a unique VLAN ID, each stack could also be configured as a primary router for its respective VLANs.

OSA also continues to support a primary and secondary router that is not associated with a specific VLAN. This primary router is referred to as the default primary router. It continues to function as the router for inbound packets that are not VLAN ID tagged, or packets that are VLAN ID tagged with a VLAN ID that is not registered to OSA. This is the same primary router support that existed prior to introduction of the VLAN ID support. Therefore, multiple specific VLAN primary routers and a single default primary router can concurrently activate and share a single OSA.

Each VLAN-specific primary and secondary router is subject to the same OSA rules (that is, supporting a single primary router and allowing multiple secondary routers) as the default primary router.

Figure 8 on page 63 shows a configuration where multiple TCP/IP stacks are sharing a single OSA, and multiple VLANs with primary routers are configured.

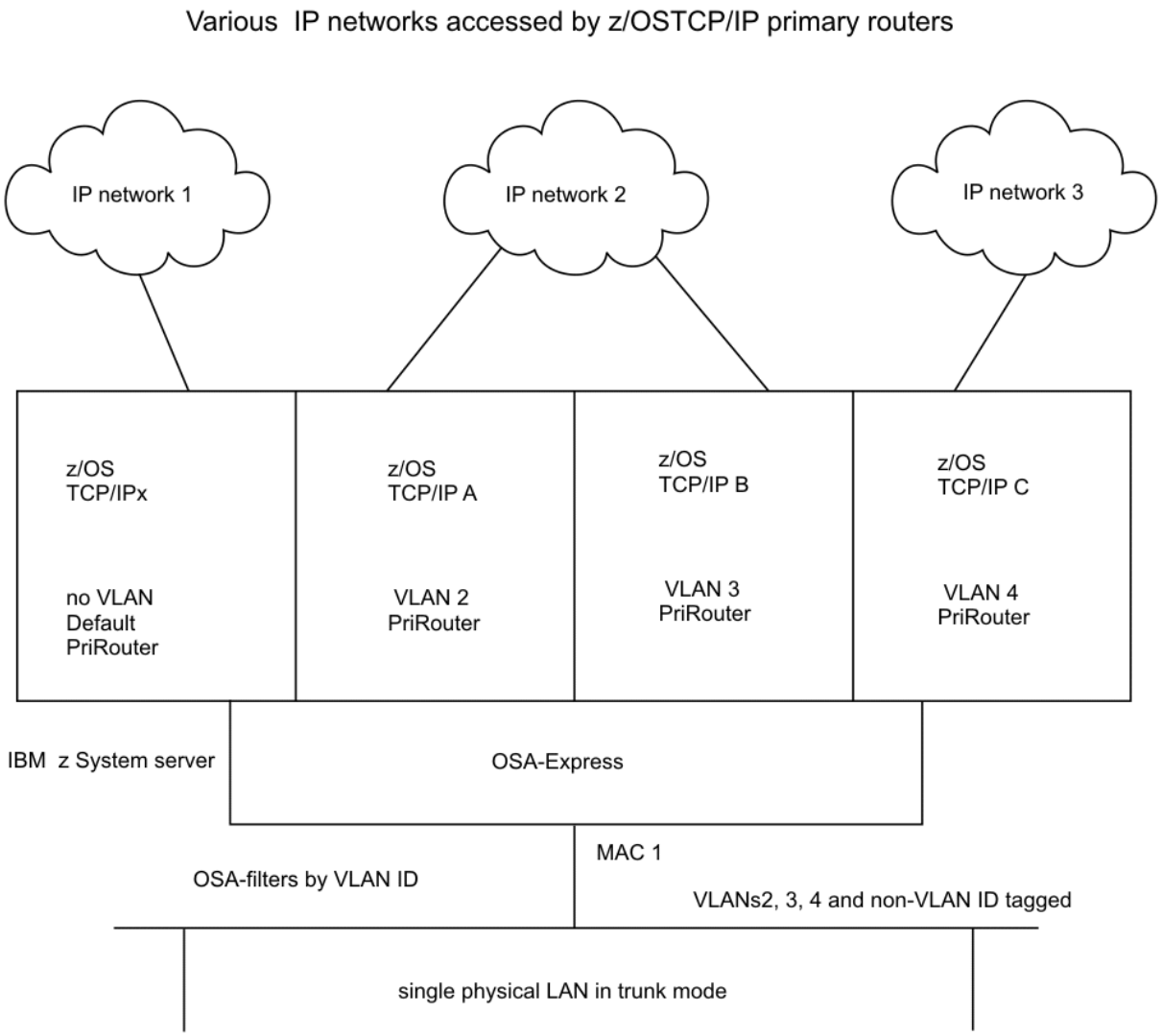


Figure 8. Primary router per VLAN (shared OSA with multiple primary routers)

In this example, TCP/IP x serves as the default primary router (PRIRouter without a VLANID configured). The other three TCP/IP stacks serve as a PRIRouter for just their specific VLANs.

For additional information regarding the details and syntax for configuring a VLAN Identifier (VLANID), and how to configure a TCP/IP stack as a primary or secondary router, see z/OS Communications Server: IP Configuration Reference. For IPv4 information on the PRIRouter, SECRouter, NONRouter, and VLANID parameters, see DEVICE and LINK – MPCIPA OSA-Express QDIO devices or INTERFACE - IPAQENET OSA-Express QDIO interfaces. For IPv6 information regarding these parameters, see INTERFACE - IPAQENET6 OSA-Express QDIO interfaces.

Network configuration strategy with VLAN

The IBM OSA-Express VLAN support allows a TCP/IP stack to register a specific VLAN identifier for both IPv4 and IPv6. Note that the VLAN ID for IPv4 can be different than the VLAN ID for IPv6. When a VLAN ID is configured, the following situations occur:

1. TCP/IP becomes VLAN aware or enabled, and this TCP/IP (IPv4 or IPv6 connection) is considered to be part of the configured VLAN.
2. During activation, TCP/IP registers the configured VLAN ID value to OSA.

3. A VLAN ID tag is added to all outbound packets.
4. OSA filters inbound VLAN ID tagged packets based on the configured VLAN ID.
5. If this stack is also configured as a router (for example, PRIRouter), the OSA primary router support is extended to this stack, allowing it to serve as the primary router for the configured VLAN ID.

When configuring a z/OS TCP/IP stack with a VLAN ID, consideration must also be given to how the LAN is partitioned and how the VLAN aware switches are configured.

VLAN switch concepts

In conjunction with the IEEE standards, most VLAN aware switches recognize and support at least two modes, referred to as trunk and access modes. This support is provided on a switch port basis. The general concepts of the two modes are as follows:

Trunk mode
Indicates that the switch should allow all VLAN ID tagged packets to pass through the switch port without altering the VLAN ID. Trunk mode is intended for servers that are VLAN capable, and filters and processes all VLAN ID tagged packets. In trunk mode, the switch expects to see VLAN ID tagged packets inbound to the switch port.

Access mode
Indicates that the switch should filter on specific VLAN IDs and allow only packets that match the configured VLAN IDs to pass through the switch port. The VLAN ID is then removed from the packet before it is sent to the server (that is, VLAN ID filtering is controlled by the switch). In access mode, the switch expects to see packets without VLAN ID tags inbound to the switch port.

For the specific details regarding how to configure a VLAN switch, consult the specific product documentation.

VLAN configuration recommendations

When deploying the z/OS TCP/IP VLAN ID support in conjunction with the IBM OSA-Express feature in QDIO mode, it is recommended that deployment be symmetrical with the configuration of the corresponding VLAN switch. Specific recommendations are as follows:

- When using a VLAN ID, configure the switch port in trunk mode.

  When a VLAN ID is configured in any z/OS TCP/IP stack that is sharing an OSA, the corresponding switch port associated with the OSA should be configured in trunk mode. In this mode, OSA performs VLAN ID filtering.

  Conversely, access mode should not be configured on the switch port if a VLAN ID is configured on any stack sharing this OSA.
- When not using a VLAN ID, configure the switch port in access mode.

  When a VLAN ID is not configured on any z/OS TCP/IP stack that is sharing an OSA, configure access mode at the switch (if VLAN filtering is wanted and therefore required at the switch).

  Conversely, trunk mode should not be configured on the switch port if a VLAN ID is not configured on any stack sharing this OSA.
- Multiple OSAs on the same physical LAN

  When a z/OS TCP/IP stack has access to multiple OSAs that are on the same physical LAN, and a VLAN ID is configured on any of the OSAs, it is recommended that this stack configure a VLAN ID for all OSAs on the same physical LAN. That is, do not mix VLAN and no-VLAN on the same physical network when a stack has access to the same LAN through multiple OSAs.
- VLAN ID 1 considerations

  Some switch vendors use VLAN ID 1 as the default value when a VLAN ID value is not explicitly configured. It is recommended that you avoid the value of 1 when configuring a VLAN ID value.

Figure 9 on page 65, Figure 10 on page 66, and Figure 11 on page 67 illustrate the preceding recommendations.

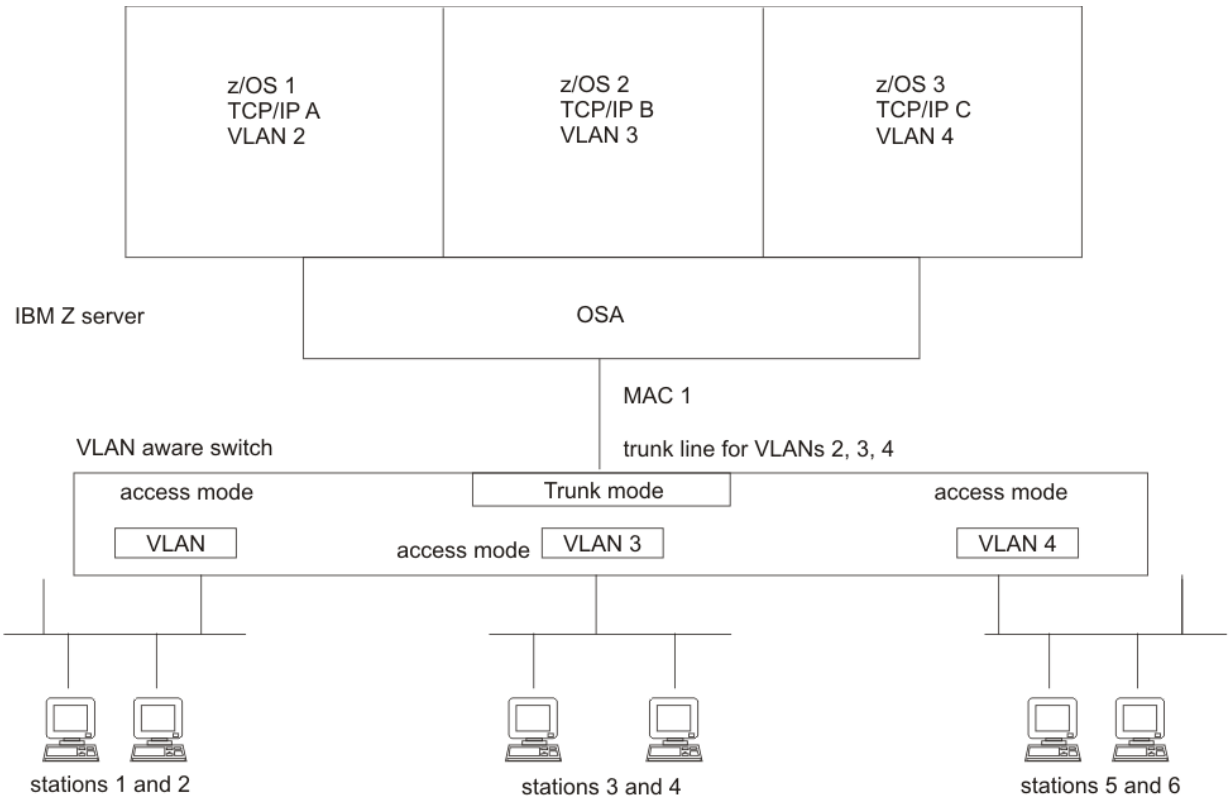


Figure 9. Single OSA and VLAN switch configuration

Figure 9 on page 65 shows the recommended VLAN switch port configuration when a VLAN ID is configured in the TCP/IP stack. A single physical LAN is divided into three separate virtual LANs (2, 3, and 4), the OSA port is configured as a trunk line, and the other ports on the switch are configured in access mode for their specific VLAN.

In Figure 9 on page 65 there are three virtual LANs deployed through the same shared OSA, where each TCP/IP stack appears to have a unique and isolated physical network as follows:

- VLAN 2 - TCP/IP A and stations 1 and 2
- VLAN 3 - TCP/IP B and stations 3 and 4
- VLAN 4 - TCP/IP C and stations 5 and 6

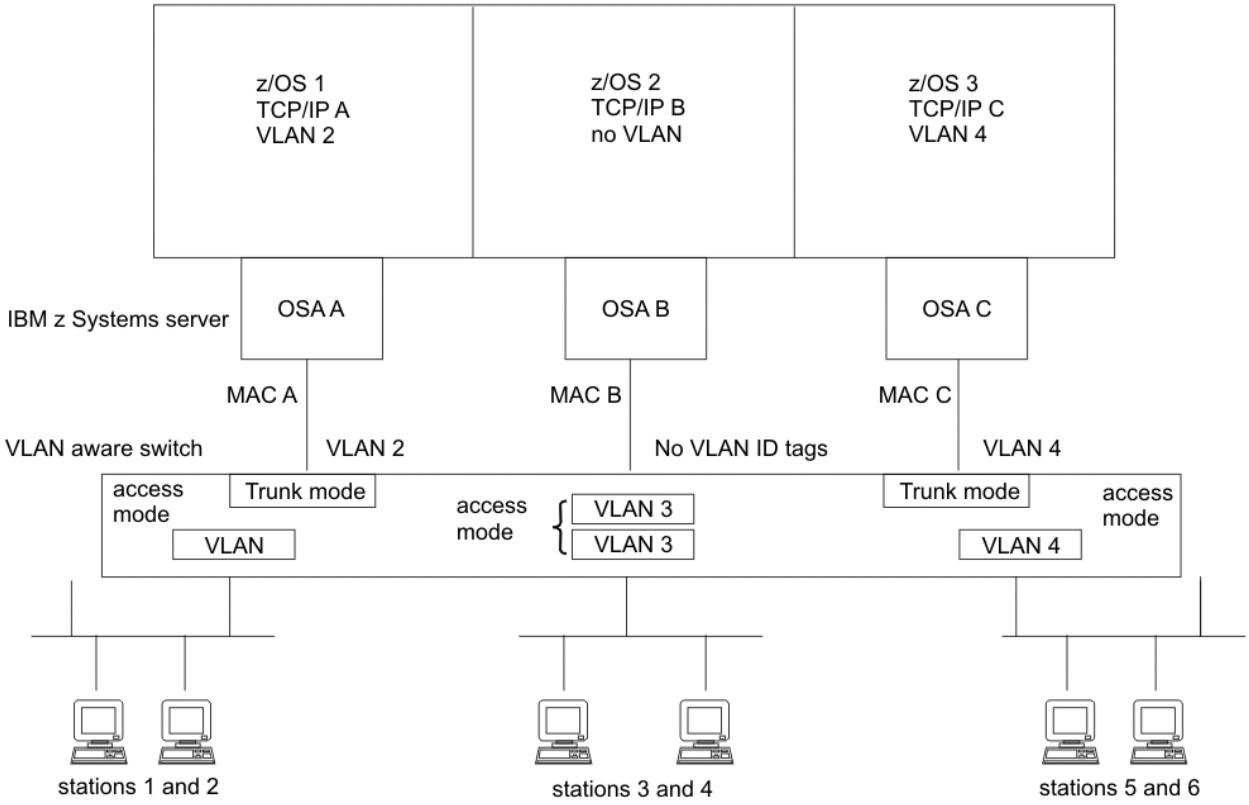


Figure 10. Matching VLAN switch configuration to multiple OSAs (VLAN configuration)

Figure 10 on page 66 illustrates using multiple OSAs and TCP/IP stacks. Three unique VLANs are created. However, TCP/IP stack B will not deploy a VLAN ID, and the corresponding switch port is configured in access mode. No VLAN ID tags will flow to this OSA port.

In Figure 10 on page 66 there are also three virtual LANs deployed. Access to each VLAN is provided through separate OSAs, yet the functionality of having three physical networks is still provided. TCP/IP B is not configured with a VLAN ID, and therefore stack B is unaware of the existence of VLAN 3 (although stations 3 and 4 on VLAN 3 have access to stack B through OSA B). Note that the switch port for OSA B is configured in access mode, while the other two switch ports are configured in trunk mode.

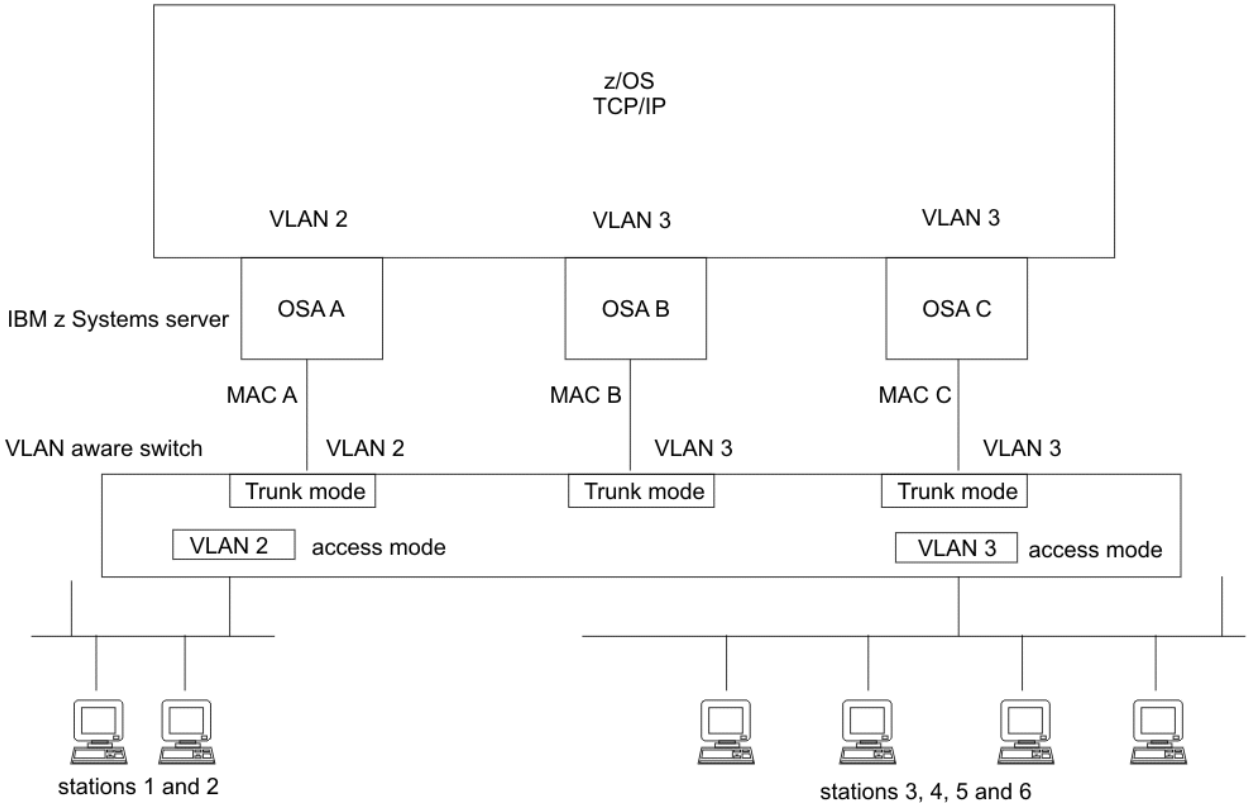


Figure 11. Single stack using multiple OSAs on the same physical network

Figure 11 on page 67 illustrates a single TCP/IP stack using multiple OSAs that are on the same physical network. There are two VLANs deployed, where OSA A is on VLAN 2, and OSA B and OSA C are on VLAN 3.

Configuring OSA B and OSA C with the same VLAN ID has significance for failure or takeover scenarios. The interface takeover (ARP takeover) function, with redundant connectivity onto a LAN, applies within the VLAN. Therefore, if OSA B becomes unavailable, OSA C can take over. Similarly, OSA B can take over if OSA C becomes unavailable. However, OSA A cannot take over for either OSA B or OSA C, because OSA A is on a different VLAN.

In Figure 11 on page 67, a single TCP/IP stack has access to two VLANs through three OSAs, which provides the following network isolation:

- VLAN 2 - through OSA A to stations 1 and 2
- VLAN 3 - through OSA B and OSA C to stations 3, 4, 5 and 6

OSA-Express port sharing

An OSA-Express port can be shared by multiple TCP/IP stacks. In such a configuration, when a unicast packet is sent over the OSA-Express port and the next-hop IP address is registered by another TCP/IP stack on the same LAN or VLAN that shares the OSA-Express port, then the OSA adapter routes the packet directly to the stack that shares the port, bypassing the external LAN. For multicast and broadcast, the OSA adapter also routes the packet directly to the stack that shares the port, in addition to sending the packet onto the LAN.

OSA-Express connection isolation

OSA-Express connection isolation provides a way to prevent the adapter from internally routing packets directly to a stack that shares the same port. When connection isolation is in effect, the OSA-Express feature discards any unicast packets when the next-hop address is registered by a stack sharing the same port, and prevents any multicast or broadcast packets from being internally routed between the stacks sharing the port.

For direct routing to occur, the OSA-Express feature requires that neither of the stacks that are sharing a port can be isolated. Therefore, for traffic between two stacks that are sharing a port, as long as at least one of the stacks is isolated, then connection isolation is in effect for traffic in both directions between these stacks.

OSA-Express connection isolation can be useful when you want to prevent communication between two stacks that share the same OSA-Express port, and it provides extra assurance against a misconfiguration that might otherwise allow such traffic to flow. OSA-Express connection isolation can also be useful if you want to ensure that traffic flowing through the OSA adapter does not bypass any security features implemented on the external LAN.

Dynamic routing is not aware of OSA-Express connection isolation, which is an issue only if static routes are not used and traffic needs to flow between the two hosts that share the OSA adapter using connection isolation. In this case, a dynamic routing protocol might choose a route between the hosts that includes connection isolation, which would make each host unreachable from the other host. If you want dynamic routing to work between hosts that are using OSA-Express connection isolation, you must ensure in your dynamic routing configuration that the path that includes connection isolation is not chosen to route between the hosts.

Guideline: You can ensure that a path that includes OSA-Express connection isolation is not chosen as the route between two hosts by assigning higher routing costs to isolated interfaces than to other network paths between those hosts (for example dynamic XCF or MPC), so that the other network paths between the hosts are chosen. You can also accomplish this by excluding the interfaces with connection isolation from the dynamic routing domain, if it is not necessary for them to be reachable from the wider network (for example, by defining them with INTERFACE statements in OMPROUTE).

Tip: If you want traffic to flow between two stacks that share an OSA-Express port but you also want to ensure that the traffic flows over an external LAN, take one of the following actions:

- Configure each stack on a separate virtual LAN (VLAN).
- Use a static route with the next-hop address of a router on the LAN.

  Result: Using a static route with the next-hop address of a router on the LAN to route to another host on the same LAN can result in excessive ICMP redirect packets from the router to the originating host.

  Guideline: If you use this technique, turn off receipt of ICMP redirects on the sharing hosts and, if possible, configure the router to not send ICMP redirects.

ARP offload and VIPA ARP processing

In QDIO mode, the OSA performs all Address Resolution Protocol (ARP) processing for IPv4. The z/OS stack informs the OSA of the IP addresses for which it should perform ARP processing. Because the z/OS stack also supports configurations in which ARPs flow for VIPAs (which you might see on some flat network configurations that use static routing), the stack also informs the OSA of the VIPAs for which it should perform ARP processing. OSA sends gratuitous ARPs for these IP addresses during interface takeover scenarios to provide fault tolerance.

If you define the OSA using DEVICE and LINK statements, then the stack informs OSA to perform ARP processing for all VIPAs that are in the home list, which can result in numerous unnecessary gratuitous ARPs for VIPAs in an interface takeover scenario. However, if you use the IPv4 INTERFACE statement for IPAQENET, you can control this VIPA ARP processing by configuring a subnet mask for the OSA. If you specify a nonzero num_mask_bits value on the IPADDR parameter of the INTERFACE statement, then the stack informs OSA to perform ARP processing for a VIPA only if the VIPA is configured in the same subnet as the OSA (as defined by the resulting subnet mask).

Checksum offload

When packets are sent or received over the OSA-Express feature in QDIO mode, TCP/IP supports offloading most IPv4 and IPv6 inbound and outbound checksum processing (IP header, TCP, and UDP checksums) to the OSA adapter. This is the default behavior.

The TCP/IP stack performs checksum processing in the following cases in which checksum processing cannot be offloaded:

- IPv6 packets sent to and from OSA-Express features that support only IPv4 checksum offload (OSA-Express3 and earlier features)
- Packets that go directly to another stack that shares the same OSA-Express port when the OSA-Express feature does not support LPAR-to-LPAR checksum offload (OSA-Express3 and earlier features)
- IPSec-encapsulated packets
- Fragmented and reassembled packets
- Outbound multicast and broadcast packets
- Outbound TCP packets that contain only a TCP header
- IPv6 packets that contain extension headers
- When multipath is in effect (unless all interfaces in the multipath group support checksum offload)

In some cases, you might want to use the NOCHECKSUMOFFLOAD parameter on the IPCONFIG and IPCONFIG6 profile statements to configure the TCP/IP stack to perform all checksum processing.

TCP segmentation offload

You can use TCP segmentation offload support to have the TCP/IP stack offload most IPv4 and IPv6 outbound TCP segmentation processing in QDIO mode to an OSA-Express feature. You can configure this function by specifying the SEGMENTATIONOFFLOAD parameter on the IPCONFIG or IPCONFIG6 profile statement. You must also offload checksum processing to the OSA-Express feature to enable segmentation offload. The TCP/IP stack performs TCP segmentation processing in the following cases in which segmentation cannot be offloaded:

- IPv6 packets sent to and from OSA-Express features that support only IPv4 segmentation offload (OSA-Express3 and earlier features)
- Packets that go directly to another stack that shares the same OSA-Express port
- IPSec-encapsulated packets
- When multipath is in effect (unless all interfaces in the multipath group support segmentation offload)
- When the SEGMENTATIONOFFLOAD parameter is not specified on the IPCONFIG or IPCONFIG6 statement

Tip: Applications that use large TCP send buffers will obtain the most benefit from TCP segmentation offload. The size of the TCP receive buffer on the other side of the TCP connection also affects the negotiated buffer size. You can control the size of these buffers using the following mechanisms:

- The TCPSENDBFRSIZE parameter on the TCPCONFIG statement sets the default TCP send buffer size for all applications.
- An application can use the SO_SNDBUF socket option to override the default TCP send buffer size.
- The TCPRCVBUFRSIZE parameter on the TCPCONFIG statement sets the default TCP receive buffer size for all applications.
- An application can use the SO_RCVBUF socket option to override the default TCP receive buffer size.

Requirement: To determine the OSA microcode level required to use TCP segmentation offload support, see the Preventive Service Planning (PSP) bucket for the appropriate device.

Dynamic LAN idle timer

z/OS Communications Server can dynamically adjust how frequently an OSA-Express2 or later feature interrupts the host for inbound traffic. By monitoring traffic patterns, the stack can adjust the interrupt-timing values to maximize throughput.

To configure an OSA-Express2 or later feature for dynamic LAN idle timer, use the INTERFACE or LINK statement with the dynamic inbound performance setting (INBPERF DYNAMIC). For information about dynamic LAN idle timer and the INBPERF parameter, see DEVICE and LINK – MPCIPA OSA-Express

QDIO devices, INTERFACE - IPAQENET OSA-Express QDIO interfaces, and INTERFACE - IPAQENET6 OSA-Express QDIO interfaces in z/OS Communications Server: IP Configuration Reference.

Restriction: The OSA-Express2 or later feature must be configured in QDIO mode.

Optimized latency mode

One way to improve the performance of an OSA-Express3 or later feature in QDIO mode for processing workloads with demanding latency requirements is to configure the OSA-Express feature to operate in optimized latency mode. Optimized latency mode optimizes interrupt processing for both inbound and outbound data, which decreases latency and can provide significant increases in throughput, particularly for high volume, interactive, non-streaming workloads.

To configure an OSA-Express3 or later feature to operate in optimized latency mode, use the INTERFACE statement with the OLM parameter. Because optimized latency mode affects both inbound and outbound interrupts, it supersedes other inbound performance settings that the INBPERF parameter sets. For more information about optimized latency mode and the OLM and INBPERF parameters on the INTERFACE statement for IPAQENET and IPAQENET6, see z/OS Communications Server: IP Configuration Reference.

Because of the operating characteristics of optimized latency mode, two other configuration changes might be required.

- For outbound traffic to gain the benefit of optimized latency mode, direct traffic to priority queues 1, 2, or 3 by using the WLMPRIORITYQ parameter on the GLOBALCONFIG statement, or by using Policy Agent and configuring a policy with the SetSubnetPrioTosMask statement.

  Although an OSA-Express feature supports multiple outbound write priority queues, outbound optimized latency mode occurs only for traffic on priority queue 1 (priority level 1). The TCP/IP stack combines all the traffic that is directed to priority queues 1, 2, and 3 into priority queue 1 for any OSA-Express3 or later feature that is operating in optimized latency mode.

  For more information about directing traffic to outbound OSA-Express priority queues by using the WLMPRIORITYQ parameter on the GLOBALCONFIG statement or using the SetSubnetPrioTosMask statement, see z/OS Communications Server: IP Configuration Reference.

  Guideline: Configure the WLMPRIORITYQ parameter with no subparameters, which assigns a default mapping of service class importance levels to OSA-Express outbound priority queues. This default mapping directs traffic that is assigned to the higher priority service class importance levels 1–4 to queues that operate in optimized latency mode, and enables the appropriate types of traffic to benefit from optimized latency mode.

  Result: If neither the WLMPRIORITYQ or SetSubnetPrioTosMask statements are specified, any packet that has a type of service (ToS) byte with the first 3 bits being 000 or 001 is directed to queue 4 and does not benefit from optimized latency mode.
- To achieve optimal latency for one or more network interfaces that are operating in optimized latency mode, limit the number of network interfaces that can concurrently share an OSA-Express3 or later feature.

  When at least one network interface is operating in optimized latency mode, ensure that no more than four concurrent network interfaces are sharing an OSA-Express3 or later port, and no more than eight concurrent network interfaces are sharing an OSA-Express3 or later channel path identifier (CHPID). The following configurations can result in multiple users that are sharing an OSA-Express3 or later feature.

  – Multiple LPARs sharing the OSA-Express3 or later feature
  – Multiple stacks on the same LPAR sharing the OSA-Express3 or later feature
  – Multiple VLAN interfaces to the OSA-Express3 or later feature
  – Both an IPv4 and an IPv6 active interface to the OSA-Express3 or later feature
  – A TCP/IP stack that is enabling the OSA-Express network traffic analyzer (OSAENTA) for the OSA-Express3 or later feature

Optimized latency mode is intended for high volume, interactive workloads. Although optimized latency mode can compensate for some mixing of workloads, an excessive amount of high volume, streaming workloads, such as bulk data or file transfer, can result in higher processor consumption.

Guideline: When multipath routing by using the PERPACKET option is enabled, do not configure a multipath group that contains an OSA-Express3 or later feature configured with optimized latency mode and any other type of device.

Restrictions:

- Optimized latency mode is limited to OSA-Express3 or later Ethernet features in QDIO mode that are running with an IBM System z10® or later server. For more information, see the 2097DEVICE Preventive Service Planning (PSP) bucket.
- Traffic that is either inbound over or being forwarded to an OSA-Express3 or later feature configured to operate in optimized latency mode is not eligible for the accelerated routing that HiperSockets Accelerator and QDIO Accelerator provide.
- For an OSA-Express3 or later interface configured to operate in optimized latency mode, the stack ignores the configured INBPERF setting and uses the value DYNAMIC.

QDIO inbound workload queueing

OSA-Express3 or later features can perform a degree of traffic sorting by placing inbound packets for differing workload types on separate processing queues. This function is called QDIO inbound workload queueing (IWQ). With the inbound traffic stream already sorted by the OSA-Express feature, z/OS Communications Server provides the following performance optimizations:

- Finer tuning of read-side interrupt frequency to match the latency demands of the various workloads that are serviced
- Improved multiprocessor scalability, because the multiple OSA-Express input queues are now efficiently serviced in parallel

When QDIO IWQ is enabled, z/OS Communications Server and the OSA-Express feature establish a primary input queue and one or more ancillary input queues (AIQs), each with a unique read queue identifier (QID) for inbound traffic. z/OS Communications Server and the OSA-Express feature cooperatively use the multiple queues in the following way:

- The OSA-Express feature directs an inbound packet (received on this interface) that is to be forwarded by the sysplex distributor to the sysplex distributor AIQ. z/OS Communications Server then tailors its processing for the sysplex distributor queue, notably by using the multiprocessor to service sysplex distributor traffic in parallel with traffic on the other queues.
- The TCP layer automatically detects connections operating in a bulk-data fashion (such as the FTP data connection), and these connections are registered to the receiving OSA-Express feature as bulk-mode connections. The OSA-Express feature then directs an inbound packet (received on this interface) for any registered bulk-mode connection to the TCP bulk-data AIQ. z/OS Communications Server tailors its processing for the bulk queue, notably by improving in-order packet delivery on multiprocessors, which likely results in improvements to CPU consumption and throughput. Like other AIQs, processing for data on the bulk queue can be in parallel with traffic on the other queues.
- The OSA-Express feature directs an inbound Enterprise Extender packet (received on this interface) to the Enterprise Extender AIQ. This allows z/OS Communications Server to process inbound traffic on the Enterprise Extender queue in parallel with inbound traffic on the other queues for this interface.
- The OSA-Express feature directs inbound AH packets, ESP packets, and UDP-encapsulated ESP packets that are received on this interface to the IPSec AIQ. This allows z/OS Communications Server to process inbound traffic on the IPSec queue in parallel with inbound traffic on the other queues for this interface.
- The OSA-Express feature directs inbound packets (received on this interface) that are destined to a zCX server to the zCX AIQ. This allows z/OS Communications Server to process inbound traffic on the zCX queue in parallel with inbound traffic on the other queues for this interface. Additionally, it also allows for this inbound traffic to be processed on a System z Integrated Information Processor (zIIP).

- If a packet is not directed to an AIQ, the OSA-Express feature directs the packet to the primary input queue.

Requirement: QDIO IWQ is limited to OSA-Express3 or later Ethernet features in QDIO mode that are running on an IBM System z10 GA3 or later server. For more information, see the 2097DEVICE, 2098DEVICE, or 2817DEVICE Preventive Service Planning (PSP) buckets as appropriate.

Requirement: QDIO IWQ for IPSec and zCX is limited to OSA-Express6S or later Ethernet features in QDIO mode that are running on an IBM z14® or later server. For more information, see the 3906DEVICE or 3907DEVICE Preventive Service Planning (PSP) bucket.

Results:

- If you already enabled IWQ on an OSA-Express6S, installed the IWQ for IPSec enablement PTF (TCP/IP APAR PI77649), and have IPSec enabled, z/OS Communications Server will automatically define the IPSec input queue. This new queue will not be backed by CSM storage until the first IPSec tunnel is activated on that stack. Therefore, if you activate any IPSec tunnels, each OSA-Express6S interface on that stack that enables IWQ will consume an additional 36 K of fixed ECSA and an additional amount of fixed 64-bit CSM equal to the READSTORAGE setting for the interface.
- If you already enabled IWQ on an OSA-Express6S and installed the IWQ for zCX enablement PTFs (VTAM APAR OA58300 and TCP/IP APAR PH16581), z/OS Communications Server will automatically define the zCX input queue. This new queue will not be backed by CSM storage until the first zCX dynamic VIPA is activated on that stack. Therefore, if you activate any zCX dynamic VIPA, each OSA-Express6S interface on that stack that enables IWQ will consume an additional 36 K of fixed ECSA and an additional amount of fixed 64-bit CSM equal to the READSTORAGE setting for the interface.

Restrictions:

- QDIO IWQ is not supported for IPAQENET interfaces defined with the DEVICE, LINK, and HOME statements. You must convert your IPAQENET definitions to use the INTERFACE statement to enable this support. For more information, see "Steps for converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the IPv4 IPAQENET INTERFACE statement" on page 57.
- QDIO IWQ is not supported for a z/OS guest on z/VM® using simulated (virtual) devices, such as virtual switch (VSWITCH) or guest LAN.
- Bulk-mode TCP connection registration is supported only in configurations in which a single inbound interface is servicing the bulk-mode TCP connection. If a bulk-mode TCP connection detects that it is receiving data over multiple interfaces, QDIO IWQ is disabled for the TCP connection and inbound data from that point forward is delivered to the primary input queue.
- QDIO IWQ does not apply for traffic that is sent over an OSA port that is shared by the receiving TCP/IP stack when an indirect route (where the next hop and destination IP address are different) is being used; this traffic is placed on the primary input queue. QDIO IWQ does apply when traffic on the shared OSA path uses a direct route (where the next hop and destination IP address are the same).

Guideline: The WORKLOADQ parameter requires an additional amount of fixed 4K CSM HVCOMMON storage per AIQ. The amount of storage consumed per AIQ is based on the amount of storage defined for READSTORAGE for this interface. The bulk AIQ is always backed with this additional CSM storage. The remaining AIQs are not backed by the additional CSM storage until the specific function (EE, SD, or IPSec) is used. The EE AIQ is backed by fixed 4K CSM DSPACE64 storage (instead of HVCOMMON). To verify the amount of CSM storage that is being used for each input queue, display the VTAM TRLE name that is associated with the interface. The WORKLOADQ parameter also requires an additional 36K of ECSA per AIQ.

Tip: The additional CSM storage consumed by each OSA interface using WORKLOADQ consumes fixed (real) storage. It is recommended that you verify that the additional fixed storage required by enabling WORKLOADQ (per OSA interface) will not approach any of the following limits:

- The CSM FIXED MAXIMUM value used by Communications Server (use the D NET, CSM command and see the CSM FIXED MAXIMUM value defined in IVTPRM00)
- The actual real storage available to this z/OS system (see D M=STOR or D M=HIGH)

Guideline: CPU might increase significantly on a busy Sysplex Distributor if QDIO Accelerator is not installed when using IWQ.

Steps for enabling QDIO inbound workload queueing

Use QDIO inbound workload queueing (IWQ) to establish a primary input queue and one or more ancillary input queues (AIQs). QDIO IWQ provides performance optimizations, such as the processing of data on one queue in parallel with traffic on the other queues.

Procedure

Perform the following steps to enable QDIO IWQ:

1. Convert IPAQENET definitions from DEVICE, LINK, and HOME statements to the INTERFACE statement. For more information, see "Steps for converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the IPv4 IPAQENET INTERFACE statement" on page 57.
2. Specify the INBPERF parameter with the DYNAMIC setting on the INTERFACE statement for the IPAQENET or IPAQENET6 interface. In addition, you must specify the WORKLOADQ subparameter and the VMAC parameter.

   Tip: In addition to enabling QDIO IWQ, the INBPERF DYNAMIC setting also dynamically tunes the read-side interrupt frequency for the OSA-Express feature.

Results

You know you are done when the `WorkloadQueueing` field of the Netstat DEvlinks/-d report is set to `Yes`, indicating that QDIO IWQ is enabled. You can also obtain this information with a GetIfs request for the TCP/IP callable NMI (EZBNMIFR).

You can use the Netstat ALL/-A report to determine whether any TCP connections are registered to the TCP bulk data ancillary input queue (AIQ); the `Ancillary Input Queue` field is set to `Yes` and the `BulkDataIntfName` field is set to the interface name. You can also obtain this information with a GetConnectionDetail request for the TCP/IP callable NMI (EZBNMIFR).

You can use the Netstat STATS/-S report to display the total number of TCP segments that are processed on the TCP bulk data AIQ; this number is displayed in the `Segments Received on OSA Bulk Queues` field. You can also obtain this information with a GetGlobalStats request for the TCP/IP callable NMI (EZBNMIFR).

For more information about the Netstat command, see z/OS Communications Server: IP System Administrator's Commands. For more information about EZBNMIFR, see z/OS Communications Server: IP Programmer's Guide and Reference.

Fixed storage considerations for OSA-Express interfaces in QDIO mode

OSA-Express7S with 25 GbE bandwidth

The OSA-Express7S feature has 25 GbE bandwidth compared to a maximum of 10 GbE for earlier generation OSAs. The z/OS Communications Server default value for read storage in previous releases is insufficient to accommodate a high-volume workload over a 25 GbE adapter and could result in packet loss and sub-optimal performance.

For OSA-Express interfaces with at least 25 GbE of bandwidth, z/OS Communications Server uses higher read storage values for the keywords MAX, MIN, and AVG and a higher default value than for interfaces with less than 25 GbE of bandwidth. For details on the read storage values used for the various keywords, see QDIOSTG start option in z/OS Communications Server: SNA Resource Definition Reference.

Obtaining 8 MB of fixed read storage

The largest amount of read storage that can be allocated for an OSA-Express QDIO interface is 8 MB.

Guideline: If you expect a high-volume workload over an interface with at least 25 GbE of bandwidth, you should run with 8 MB of read storage to achieve optimal performance.

You can obtain 8 MB of fixed CSM 4K HVCOMMON of read storage for an OSA-Express QDIO interface which has at least 25 GbE of bandwidth in one of the following ways:

- Use the default QDIOSTG value of MAX in the VTAM start options and use the default READSTORAGE value of GLOBAL on the INTERFACE or LINK statement for an OSA with 25 GbE bandwidth.
- Specify a READSTORAGE value of MAX on the INTERFACE or LINK statement for any OSA. This will override the global QDIOSTG setting for that OSA.
- Specify a VTAM start option of QDIOSTG=126 and specify (or default) the READSTORAGE value on the OSA INTERFACE statement to GLOBAL. This will make 8 MB the default for all OSAs. To get a lower value for a specific OSA interface, you will need to configure a value other than GLOBAL on the READSTORAGE parameter on the INTERFACE or LINK statement for that OSA.

Tip: To see the amount of read storage in MB being used for a specific QDIO interface, use the **Netstat DEvlinks/-d** command.

Additional fixed storage for OSA interfaces using 8 MB of read storage

For OSA interfaces which have a bandwidth of at least 10 GbE and which are using 8 MB of read storage, z/OS Communications Server allocates additional fixed CSM 4K HVCOMMON storage for work element processing. If the interface is not using IWQ, the additional storage is 96 KB (plus another 32 KB if QDIO Accelerator is enabled). If the interface is using IWQ, the additional storage is 224 KB (plus another 96 KB if QDIO Accelerator is enabled).

Fixed maximum storage for CSM buffers

The IVTPRM00 parmlib member defines parameters for CSM. The default value of the FIXED MAX setting is 512 MB. (In releases prior to z/OS V2R4 Communications Server, the default value was 200 MB.) This setting defines the maximum amount of storage that is dedicated to fixed CSM buffers.

Tip: If you use OSA-Express interfaces with at least 25 GbE of bandwidth and are using a FIXED MAX setting lower than the default value, you should consider increasing the FIXED MAX setting in your IVTPRM00 parmlib member to 512 MB. If you increase this setting, you should also review your HVCOMMON setting in IEASYSxx and determine whether you need to increase this value to account for the fact that the system reserves a larger portion of this storage for CSM buffers.

OSA configuration recommendations

If your configuration includes OSA-Express QDIO interfaces with different speeds (for example 10 GbE and 25 GbE), you should consider how you plan to use these interfaces in your network configuration.

Guideline: In general, you should avoid mixing interfaces with different speeds (especially when you expect a heavy workload over these interfaces) in the following configurations:

- In the same multipath group
- In the same LAN group
- In a sysplex distributor environment where traffic reaches the distributor stack over a higher bandwidth OSA and is distributed to the target stack over a lower bandwidth OSA

Tip: You can display the speed of an active OSA-Express QDIO interface by using either **Netstat DEvlinks/-d** or the **DISPLAY TCPIP,,OSAINFO** command.

Displaying OSA-Express QDIO interface information

You can use the following commands to display information about OSA-Express QDIO interfaces:

- Netstat DEvlinks/-d

This command displays configured and runtime information from the TCP/IP stack for an OSA-Express QDIO interface.

- DISPLAY TCPIP,,OSAINFO

  This command displays configured and runtime information from the OSA-Express QDIO feature for the data subchannel device of the interface.

For information about the syntax and output of these commands, see z/OS Communications Server: IP System Administrator's Commands.

HiperSockets concepts and connectivity

HiperSockets is a System z hardware feature that provides high performance internal communications between LPARs within the same central processor complex (CPC), without the use of any additional or external hardware equipment (for example, channel adapters and LANs). When the processor supports HiperSockets and the CHPIDs are configured in HCD (IOCP), TCP/IP connectivity can occur for two reasons:

- DYNAMICXCF is configured on the IPCONFIG (IPv4) or the IPCONFIG6 (IPv6) statements.
- A user-defined HiperSockets (MPCIPA) DEVICE and LINK (IPv4) or INTERFACE (IPv4 or IPv6) is configured and started.

Therefore, there are two types of HiperSockets devices:

- DYNAMICXCF HiperSockets device or interface (TRLE IUTIQDIO and an MPC group of subchannel devices)

  The PORTNAME is IUTIQD*xx*, where *xx* is the IQD CHPID that VTAM uses (for example, IUTIQDFD when using IQD CHPID x'FD').
- A user-defined HiperSockets device or interface consisting of an MPC group of subchannel devices and one of the following definitions:
  - TRLE IUTIQD*xx* for INTERFACE (IPv6) and DEVICE/LINK (IPv4)
  - TRLE IUTIQ4*xx* for INTERFACE (IPv4)

  The PORTNAME is not applicable for these TRLEs.

When HiperSockets is configured with a DEVICE and LINK definition, the z/OS TCP/IP stack has only one connection to the HiperSockets CHPID. You define this connection with a combination of the DEVICE, LINK, and HOME statements for IPv4, an INTERFACE statement for IPv6, or both. This single connection uses one DATAPATH device from the TRLE definition that VTAM creates to represent this HiperSockets CHPID. Both IPv4 and IPv6 traffic share this one connection and DATAPATH device. For more information about DATAPATH devices and TRLEs, see z/OS Communications Server: SNA Network Implementation Guide.

If you configure a HiperSockets CHPID for IPv4 using the INTERFACE statement and also configure the same HiperSockets CHPID for IPv6, then the z/OS TCP/IP stack has two connections to the OSA-Express feature and requires two DATAPATH devices.

Steps for converting from IPv4 IPAQIDIO DEVICE, LINK, and HOME definitions to the IPv4 IPAQIDIO INTERFACE statement

Use the INTERFACE statement to configure IPv4 definitions for HiperSockets, rather than using the DEVICE, LINK, and HOME statements. The INTERFACE statement simplifies stack configuration for IPAQIDIO interfaces, and some functions like multiple VLAN support require that the HiperSockets interface is defined with the INTERFACE statement.

Procedure

Perform the following steps to convert your TCP/IP profile so that it uses the INTERFACE statement to configure IPv4 definitions for HiperSockets.

1. Convert the IPv4 IPAQIDIO DEVICE, LINK, and HOME statements to an IPv4 IPAQIDIO INTERFACE statement.

   The values used in the following substeps are based on the sample profiles shown in Figure 12 on page 77 and Figure 13 on page 77.

   a. Copy the LINK name (HSINTF1 in the example) and specify this name as the interface name on the INTERFACE statement.

      Tip: When you use the original LINK name as the new INTERFACE name, you do not have to change the static route definitions, OMPROUTE definitions, PKTTRACE statement, and PRIMARYINTERFACE statement.

   b. Copy the IPAQIDIO parameter from the LINK statement and specify this parameter after the DEFINE parameter of the INTERFACE statement.

   c. Copy the remaining LINK parameters and values (IPBCAST in the example) to the INTERFACE statement.

   d. Copy the *xx* portion of the DEVICE name (FE in the example) and specify this value on the CHPID parameter of the INTERFACE statement.

   e. Copy the HOME list entry IP address (200.16.1.1 in the example) and specify this address on the IPADDR parameter of the INTERFACE statement.

   f. Append a value (/24 in the example) to the end of the IP address to define the subnet mask.

   g. Remove the HOME list entry.

   Tip: Optionally, you can dump the TCP/IP address space and use the CONVERT parameter on the TCPIPCS PROFILE subcommand to display the configuration information at the time of the dump. The resulting output reflects your IPAQENET, IPAQIDIO, and VIRTUAL DEVICE, LINK, and HOME definitions in INTERFACE statement format, which might be helpful in converting your profile to use INTERFACE statements. Review the output before you implement any changes. For more information about using the CONVERT parameter on the TCPIPCS PROFILE subcommand, see z/OS Communications Server: IP Diagnosis Guide.

   For more information about the IPv4 IPAQIDIO INTERFACE statement, see z/OS Communications Server: IP Configuration Reference.
2. Remove any BSDROUTINGPARMS entries for the interface.
3. If you configured the SOURCEVIPA parameter on the IPCONFIG statement, perform the following steps:

   a) Find the IPAQIDIO LINK in the original HOME list and search backwards to locate the static VIPA (if any) that is located closest to this link in the HOME list.

   b) If you find a static VIPA, add the SOURCEVIPAINTERFACE parameter to the IPv4 INTERFACE statement. Use the static VIPA link name as the SOURCEVIPAINTERFACE value.
4. If you are using the START statement to start the IPv4 device, change the START statement to specify the name of the IPv4 INTERFACE.
5. Ensure that your HCD configuration is sufficient for the resulting dynamic TRLE definitions. DEVICE/LINK and IPv6 INTERFACE definitions use the IUTIQDxx TRLE and the new IPv4 INTERFACE definitions use the IUTIQ4xx TRLE where xx is the HiperSockets CHPID. If you use a combination of these definitions for a HiperSockets CHPID that requires both TRLEs, configure an additional set of 10 IQD devices in HCD for the new TRLE. If you are converting from DEVICE/LINK to IPv4 INTERFACE and are not using IPv6, you can reuse the same set of IQD devices for the new TRLE. However, if VTAM is still active from your previous definitions such that the IUTIQDxx TRLE was created, recycle VTAM to reuse that set of IQD devices for the new IUTIQ4xx TRLE. For more information about dynamic TRLEs, see Resources automatically activated by VTAM in z/OS Communications Server: SNA Network Implementation Guide.

Results

For information about the TCP/IP profile (PROFILE.TCPIP) and configuration statements, see z/OS Communications Server: IP Configuration Reference.

The following examples show some additional changes that you might need to make to the definitions for HiperSockets.

```
IPCONFIG
       SOURCEVIPA
     ;
     DEVICE VIPA4811  VIRTUAL  0
     LINK   VIPA4811L VIRTUAL  0    VIPA4811
     ;
     DEVICE IUTIQDFE  MPCIPA
     LINK   HSINTF1       IPAQIDIO      IUTIQDFE
       IPBCAST
     ;
     HOME
       10.81.1.1      VIPA4811L
       200.16.1.1     HSINTF1
     ;
     PRIMARYINTERFACE HSINTF1
     ;
     BSDROUTINGPARMS TRUE
       VIPA4811L 1492 0 255.255.255.0 0
       HSINTF1   1492 0 255.255.255.0 0
     ENDBSDROUTINGPARMS
     ;
     BEGINROUTES
       ROUTE 200.16.0.0/16  =  HSINTF1  MTU 1492
     ENDROUTES
     ;
     START IUTIQDFE
```

Figure 12. Example IPv4 IPAQIDIO profile before conversion

```
IPCONFIG
       SOURCEVIPA
     ;
     DEVICE VIPA4811  VIRTUAL  0
     LINK   VIPA4811L VIRTUAL  0    VIPA4811
     ;
     ; Converted INTERFACE statement
     ;
     ;   - HSINTF1 is from the LINK statement
     ;   - DEFINE IPAQIDIO is from the LINK statement
     ;   - IPBCAST is from the LINK statement
     ;   - IPADDR 200.16.1.1 is from the HOME list entry, /24 is from the
     ;     BSDROUTINGPARMS entry subnet mask
     ;   - CHPID FE is from the DEVICE statement
     ;   - SOURCEVIPAINTERFACE VIPA4811L is from the order of the HOME list
     ;     entries
     ;
     INTERFACE HSINTF1
       DEFINE IPAQIDIO
       IPADDR 200.16.1.1/24
       CHPID   FE
       SOURCEVIPAINTERFACE VIPA4811L
     ;
     ; HSINTF1 is removed from the HOME list
     ;
     HOME
       10.81.1.1      VIPA4811L
     ;
     PRIMARYINTERFACE HSINTF1
     ;
     ; HSINTF1 is removed from BSDROUTINGPARMS
     ;
     BSDROUTINGPARMS TRUE
       VIPA4811L 1492 0 255.255.255.0 0
     ENDBSDROUTINGPARMS
     ;
     BEGINROUTES
       ROUTE 200.16.0.0/16  =  HSINTF1  MTU 1492
     ENDROUTES
     ;
     ; START statement uses the interface name
     ;
     START HSINTF1
```

Figure 13. Example IPv4 IPAQIDIO profile after conversion

Concepts and considerations for the IQD CHPID

The HiperSockets hardware device is represented by the IQD CHPID and its associated subchannel devices. All LPARs that are configured (HCD) to use the same IQD CHPID have internal connectivity and therefore have the capability to communicate using HiperSockets. If the system supports multiple channel subsystems, and if HiperSockets connectivity is required across multiple channel subsystems, the IQD CHPID must also be configured (HCD) to span the applicable channel subsystems. The IQD CHPID can be viewed as a logical LAN within the CPC. The HiperSockets hardware allows up to four (16 with systems that support multiple channel subsystems) separate IQD CHPIDs to be defined per CPC, creating the capability of having four (16 with systems that support multiple channel subsystems) separate logical LANs within the same CPC. Figure 14 on page 78 and Figure 15 on page 79 illustrate this concept:

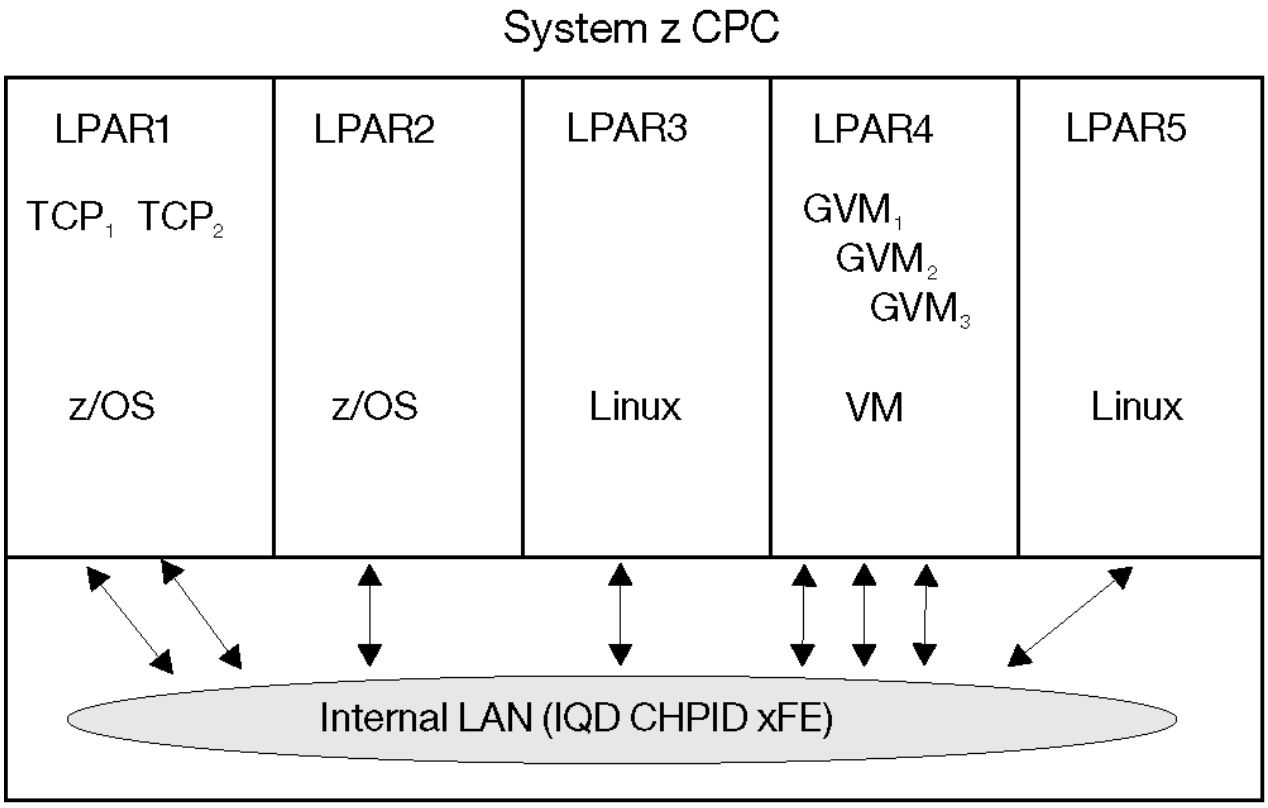


Figure 14. HiperSockets internal LAN

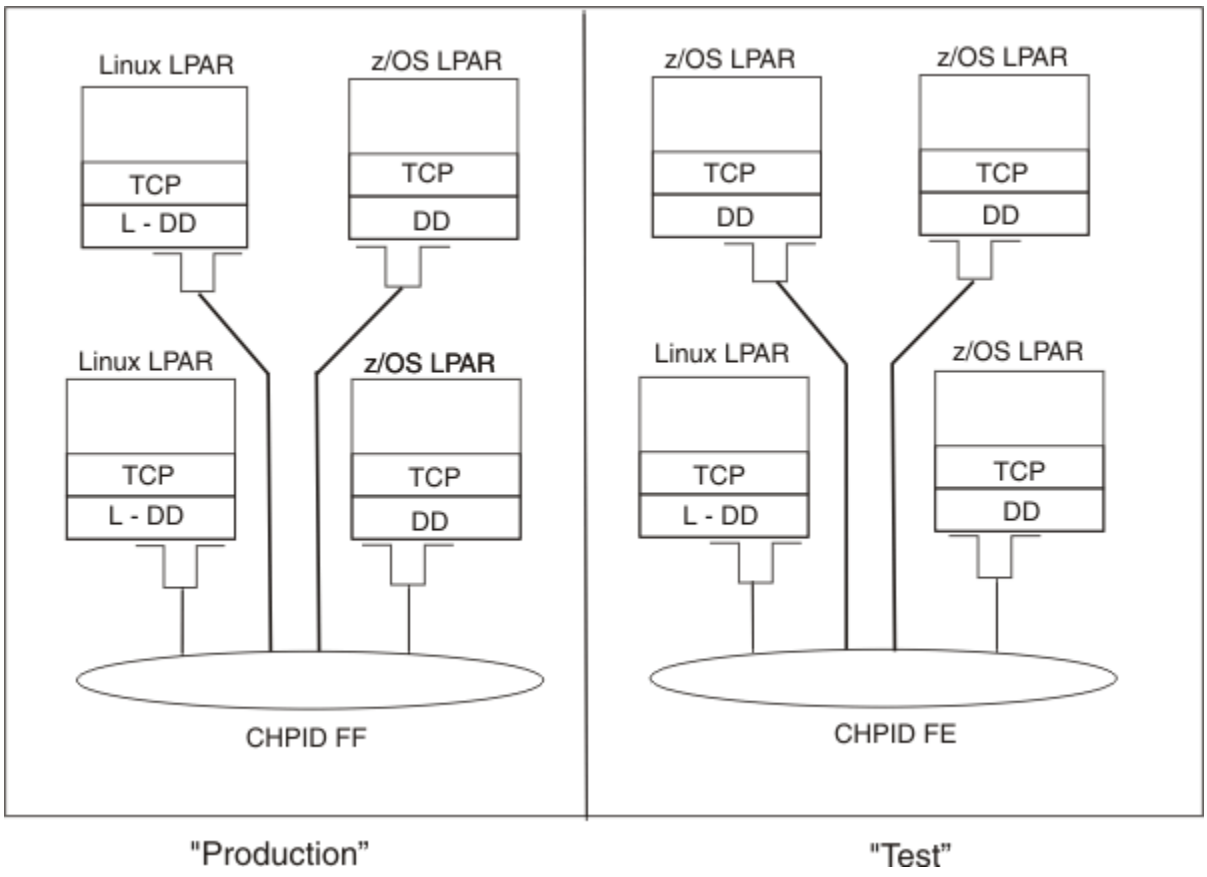


Figure 15. HiperSockets multiple internal LANs

With this capability, the system administrator can logically separate (or control) the internal connectivity. This is accomplished by controlling which specific LPARs are allowed to internally connect using HiperSockets. Further examples of this are as follows:

- SYSPLEX 'A' LPARs running on LPs 1 through 4 could use IQD CHPID x'FC'.
- SYSPLEX 'B' LPARs running on LPs 5 through 8 could use IQD CHPID x'FD'.
- A VM LPAR runs in LP 9 running various second level systems (Linux and z/OS) which use IQD CHPID x'FE'.
- Combinations of these examples could be:
  - Another set of LPARs on LPs 10 through 12 which are not using DYNAMICXCF (non SYSPLEX) are connected to IQD CHPIDs x'FE' and x'FF'.
  - Subsets of LPARs 1 through 8 are using both the DYNAMICXCF IQD CHPIDs and a non-DYNAMICXCF IQD CHPIDs.
  - Some LPARs are connected to all four IQD CHPIDs.

In a HiperSockets CHPID, IP addresses of the HiperSockets interfaces can be in the same subnet or in different subnets. Some applications can detect when HiperSockets IP addresses are in different logical subnets and might issue warning or error messages. For further subdivision of connectivity in the HiperSockets CHPID using virtual LANs, see “HiperSockets and VLAN” on page 79.

HiperSockets and VLAN

HiperSockets supports VLANs, which means you can logically subdivide the internal LAN for a HiperSockets CHPID into multiple virtual LANs. Therefore, two stacks that configure the same VLAN ID for the same CHPID can communicate over HiperSockets, while two stacks that configure different VLAN IDs cannot. For HiperSockets interfaces that share a single DATAPATH device, the VLAN ID provided applies to both IPv4 and IPv6 connections.

Guideline: If you configure a VLAN ID on a stack using a HiperSockets CHPID, configure a VLAN ID for all other stacks using the same CHPID.

When you use VLAN IDs, the z/OS TCP/IP stack can have multiple connections to the same HiperSockets CHPID. One connection is allowed for each unique combination of VLAN ID and IP version (IPv4 or IPv6). Each connection is defined by an INTERFACE statement and uses one DATAPATH device from the TRLE definition created by VTAM. To configure one or more VLANs for a single HiperSockets CHPID, see "Steps for configuring virtual LANs for a HiperSockets CHPID" on page 80.

Restrictions:

- A stack can support a maximum of 8 VLAN interfaces per IP version to the same HiperSockets CHPID.
- If multiple stacks on the same z/OS image are using the same HiperSockets CHPID, then those stacks can collectively activate a maximum of 8 VLAN interfaces per IP version for that CHPID.

Steps for configuring virtual LANs for a HiperSockets CHPID

You can logically subdivide the internal LAN for a HiperSockets CHPID into multiple virtual LANs (VLANs).

Procedure

Perform the following steps to configure one or more VLANs for a single HiperSockets CHPID:

1. Configure the interfaces.

   Configure each IPv4 interface for this HiperSockets CHPID in the TCP/IP profile using the INTERFACE statement for IPAQIDIO, rather than the DEVICE, LINK, and HOME statements.

   Configure each IPv6 interface for this HiperSockets CHPID in the TCP/IP profile using the INTERFACE statement for IPAQIDIO6.
2. Configure a VLANID value on each IPv4 INTERFACE statement and each IPv6 INTERFACE statement on this stack for this HiperSockets CHPID.

   Within each IP version, these VLANID values must be unique.
3. Configure a unique subnet for each IPv4 interface for this HiperSockets CHPID, using the subnet mask specification on the IPADDR parameter on the INTERFACE statement.
4. If you are using OMPROUTE and OMPROUTE is not configured to ignore this interface, ensure that the subnet mask value that you configure on the INTERFACE statement in the TCP/IP profile matches the subnet mask that is used by OMPROUTE for this interface.

   The subnet mask that OMPROUTE uses is the subnet mask value that is defined on the corresponding OMPROUTE statement (OSPF_INTERFACE, RIP_INTERFACE, or INTERFACE) for this interface. If no OMPROUTE statement is specified for this interface, the subnet mask that OMPROUTE uses is the class mask for the interface IP address.
5. Ensure that you understand the fixed storage requirements for this configuration.

   Each interface requires its own DATAPATH device, and each DATAPATH device requires fixed storage for read processing. For more information, see "Fixed storage requirements for OSA-Express QDIO and HiperSockets interfaces" on page 106

Planning for IQD CHPID spanning

Figure 16 on page 81 illustrates how an IQD CHPID can span both logical channel subsystem 0 and logical channel subsystem 1.

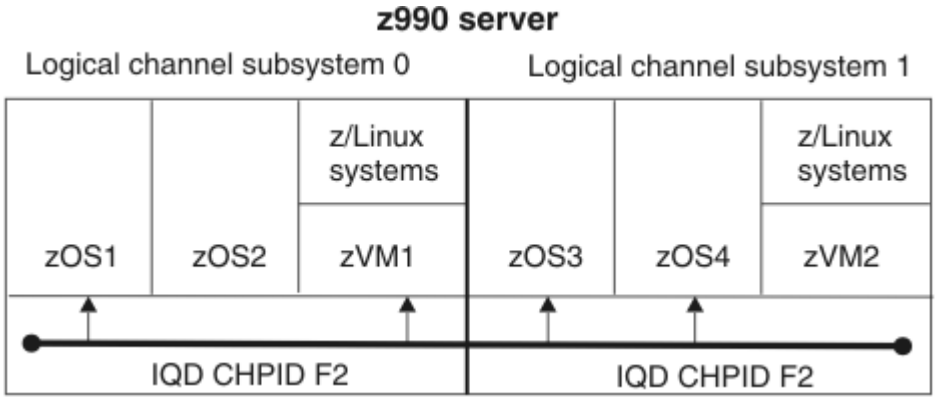


Figure 16. Spanned IQD (HiperSockets) CHPID

In Figure 16 on page 81, note:

- CHPID spanning is applicable to servers that support multiple channel subsystems, such as the IBM z990 server.
- IQD CHPIDs and their spanning attributes are configured using HCD.
- Logical channel subsystem 0 and logical channel subsystem 1 are both using IQD CHPID F2, which was configured (in HCD) to span.
- All images can exploit IQD CHPID spanning.
- Using HiperSockets support along with IQD spanning support, logical partitions in both logical channel subsystem 0 and 1 have internal connectivity.
- Specific logical partitions can be configured to have access to the spanned IQD CHPID. In this example:
  - IQD CHPID F2 is configured to span both logical channel subsystem 0 and logical channel subsystem 1.
  - In logical channel subsystem 0, zOS1 and zVM1 have connectivity (through IQD CHPID F2) to zOS3 and zOS4 in logical channel subsystem 1.
  - zOS2 and zVM2 do not have access to IQD CHPID F2.
- Up to 16 IQD CHPIDs can be configured in HCD.
- There are no TCP/IP or VTAM configuration changes required (spanning support is transparent to z/OS Communications Server).

  Tip: TCP/IP Dynamic XCF support uses HiperSockets connectivity when the remote system is reachable through HiperSockets. This is dynamically determined, and this internal logic also handles a spanned IQD CHPID.

To configure a spanned IQD CHPID, see z/OS HCD User's Guide.

The HiperSockets MPC group

For a specific IQD CHPID, VTAM builds a single HiperSockets MPC group, using the subchannel devices associated with that IQD CHPID. VTAM requests up to 10 devices from the I/O subsystem if they are defined to the CHPID being initialized, regardless of how many will be required for the number of TCP/IP stacks to be used in the LPAR. VTAM requires two subchannel devices for the read and write control devices, and 1 to 8 devices for data devices, one device for each of up to 8 TCP/IP stacks.

Therefore, to build the MPC group, there must be a minimum of 3 subchannel devices defined (within HCD) and associated with the same IQD CHPID. The maximum number of subchannel devices that VTAM will use is 10 (supporting 8 data devices or 8 TCP/IP stacks) per LPAR or MVS image. The subchannel devices must be configured for the LPAR and online prior to when the TCP/IP stack is initialized. Generally, the number of HiperSockets subchannel devices you should configure to HCD per LPAR is:

```
    2      (read / write control devices)
+   N      (where N = number of TCP/IP stacks)
```

```
                 -----------------
                     N+2     (total subchannel devices per LPAR)
Example (LPAR 1 starts two TCP/IP stacks and both stacks use HiperSockets):

         -  define 4 subchannel devices on the same IQD CHPID
         -  where 2 are used for read / write control and 2 data devices are available
```

Tip: The HCD configuration is the only way to control the number of devices that VTAM uses for HiperSockets.

The first TCP/IP stack within the LPAR to initialize dynamic XCF support (DYNAMICXCF) causes the IUTIQDIO HiperSockets MPC group to be dynamically created. The first TCP/IP stack within the LPAR to initialize a user-defined HiperSockets device (TRLE IUTIQD*xx*) causes the IUTIQD*nn* HiperSockets MPC group to be dynamically created.

Each TCP/IP stack can then start the HiperSockets device, and each stack is assigned a unique (dedicated) subchannel data device from the IUTIQDIO or IUTIQD*nn* MPC group.

Guideline: Configure the IQD CHPIDs using CHPIDs x'F0' through x'FF' (but any valid CHPID value x'00' through x'FF' can be configured as TYPE = IQD). See z/OS HCD Planning and Appendix D, “Using HCD,” on page 1339 for additional details.

HiperSockets maximum frame size

The HiperSockets hardware supports four different frame sizes referred to as the HiperSockets MFS (maximum frame size). Using HCD (or IOCP), the HiperSockets MFS is configured on the IQD CHPID using the 'OS=' parameter. All LPARs communicating over the same IQD CHPID will then use the same IQD MFS. The MFS affects the largest packet that TCP/IP can transmit. TCP/IP will adjust the MTU (Maximum Transmission Unit) based on the MFS, which is discovered during activation.

The following table depicts the four possible TCP/IP MTU sizes resulting from the HiperSockets frame sizes:

OS=*value*	**HiperSockets frame size**	**TCP/IP MTU size**
00 (default)	16 KB	8 KB
40	24 KB	16 KB
80	40 KB	32 KB
C0	64 KB	56 KB

The default HiperSockets MFS is 16 KB. However, in cases in which increased bandwidth is required (such as large file transfers, file backup), a larger MFS could be used. In most workload environments the default size will result in better storage and CPU usage.

```
************************************************************
* OS  values are '00'=16K, '40'=24K, '80'=40K  and 'C0'=64K.  *
*                                                          *
* Need at least 3 addresses per z/OS, maximum of 10:       *
*   - 2 addresses for control                              *
*   - 1 address for data for each TCP stack (between 1 and 8) *
************************************************************
 ID SYSTEM=(2064,1)
*
 CHPID PATH=FC,TYPE=IQD,SHARED,OS=00
 CHPID PATH=FD,TYPE=IQD,SHARED,OS=40
 CHPID PATH=FE,TYPE=IQD,SHARED,OS=80
 CHPID PATH=FF,TYPE=IQD,SHARED,OS=C0
*
 CNTLUNIT  CUNUMBR=FC00,PATH=FC,UNIT=IQD
 IODEVICE  ADDRESS=(2C00,16),CUNUMBR=FC00,UNIT=IQD
*
 CNTLUNIT  CUNUMBR=FD00,PATH=FD,UNIT=IQD
 IODEVICE  ADDRESS=(2C10,16),CUNUMBR=FD00,UNIT=IQD
*
 CNTLUNIT  CUNUMBR=FE00,PATH=FE,UNIT=IQD
 IODEVICE  ADDRESS=(2C20,16),CUNUMBR=FE00,UNIT=IQD
*
```

```
CNTLUNIT  CUNUMBR=FF00,PATH=FF,UNIT=IQD
IODEVICE  ADDRESS=(2C30,16),CUNUMBR=FF00,UNIT=IQD
```

See z/OS HCD Planning and Appendix D, “Using HCD,” on page 1339 for additional details.

Modifying HiperSockets connectivity [TCP/IP device and link and the VTAM HiperSockets MPC group (IUTIQDIO)]

Certain modifications can be made to the HiperSockets device (MPC group) without disrupting an active TCP/IP stack.

z/OS supports dynamic I/O for the HiperSockets CHPID and subchannel devices, allowing subchannel devices to be added or removed to or from an LPAR which has already been IPLed.

TCP/IP supports the STOP and START command for the DYNAMICXCF HiperSockets device (IUTIQDIO). However, the commands are supported only when the (internal) start (activation) was successful during stack initialization. TCP/IP also supports the STOP and START command for the user-defined HiperSockets devices (IUTIQDxx). Because a user-defined HiperSockets device is supported as an MPCIPA device, STOP and START function just as they would for other MPCIPA devices.

VTAM supports a MODIFY VTAMOPTS command that you can use to change the initial setting of the IQDCHPID start option.

Therefore, it is possible to make certain changes to the DYNAMICXCF HiperSockets MPC group (IUTIQDIO) without restarting VTAM or an active TCP/IP stack. Examples of changes that can be made are (STOP/START device required):

- Alter which specific IQD CHPID is used for DYNAMICXCF (for example, move from the x'FC' CHPID to the x'FD' CHPID).
- Add or remove subchannel devices (for example, from the current IQD CHPID).
- Alter the IQD MFS which alters the TCP/IP MTU (for example, increase the current IQD CHPID from 16k to 64k).

Although VTAM supports modifications to the start option IQDCHPID (and the modification will be immediately displayed), the effects will vary depending on what the current usage was and the change (from or to) that was made. For example:

- When MODIFIED from ANY (or CHPID) to NONE, there is no effect on current usage but blocks subsequent activations of the DYNAMICXCF HiperSockets device.
- When MODIFIED from NONE to ANY (or CHPID), there is no effect on current usage but allows subsequent activations.
- When MODIFIED from CHPID_X to CHPID_Y, there is no effect on current usage.

  Note: VTAM uses the CHPID value only when building the IUTIQDIO MPC group.

  To change CHPIDs for an active MPC group, take the following steps:

  1. TCP/IP IUTIQDIO devices and IQDIOINTF6 interfaces that are changing must be stopped.
  2. Make any necessary HCD/IOCDS changes.
  3. Verify new subchannel devices are varied online.
  4. Verify the MPC group has deactivated (with no usage it times out after approximately 2 minutes).
  5. Modify IQDCHPID = CHPID (to new CHPID).
  6. Restart the TCP/IP IUTIQDIO devices and IQDIOINTF6 interfaces.

To use HiperSockets communications, the processor must have the necessary hardware support. If the processor does not support HiperSockets communications, modifications to this start option will not be accepted, and the IQDCHPID option will not be displayed (displayed as ***NA***).

HiperSockets connectivity and routing

For each pair of stacks within a sysplex that are not on the same MVS image, if all of the following conditions are true, the stacks will use HiperSockets DYNAMICXCF connectivity (versus standard XCF connectivity):

- The two stacks must be on the same CPC.
- For the DYNAMICXCF HiperSockets device (IUTIQDIO):
  - The two stacks must be using the same IQD CHPID.
  - If running on a system that supports multiple channel subsystems and the two stacks are also in different channel subsystems, the IQD CHPID must be configured (HCD) to span.
- Both stacks must be configured (HCD) to use HiperSockets.
- For IPv4 HiperSockets connectivity, both stacks must be at the z/OS V1R2 level, or a later release. For IPv6 HiperSockets connectivity, both stacks must be at the z/OS V1R7 level.
- The initial HiperSockets activation must complete successfully.

When an IPv4 DYNAMICXCF HiperSockets device and link are created and successfully activated, a subnet route is created across the HiperSockets link. The subnet is created by using the DYNAMICXCF IP address and mask. This allows any LPAR within the same CPC to be reached, even ones that are not within the sysplex. For example, an LPAR that is running z/Linux or z/VM that does not support joining the sysplex can still be reached. The z/Linux or z/VM LPAR must define at least one IP address for the HiperSockets endpoint that is within the subnet defined by the DYNAMICXCF IP address and mask.

Similarly, when an IPv6 DYNAMICXCF HiperSockets interface is created and successfully activated, a prefix route is created across the HiperSockets interface (if prefix_route_len is specified on DYNAMICXCF). This allows any LPAR within the same CPC to be reached, even ones that are not within the sysplex. For example, an LPAR that is running z/Linux or z/VM that does not support joining the sysplex can still be reached. The z/Linux or z/VM LPAR must define at least one IP address for the HiperSockets endpoint that uses the same prefix as the DYNAMICXCF IP address.

Therefore, TCP/IP can communicate with other LPARs within the CPC over the HiperSockets connectivity created by DYNAMICXCF even when the TCP/IP in the other LPAR is not part (joins or supports) of the sysplex. You can also elect to manually configure a HiperSockets device for non-sysplex communications.

When multiple stacks are within the same LPAR that supports HiperSockets, both IUTSAMEH and HiperSockets links or interfaces will coexist. In this case, it is possible to transfer data across either link or interface. Because IUTSAMEH links or interfaces have better performance, it is better to always use them for intra-stack communication. A host route will be created by DYNAMICXCF processing across the IUTSAMEH link or interface but not across the HiperSockets link or interface. To avoid using the HiperSockets link or interface for communication within the same host, the following rules should be observed:

- Specify DYNAMICXCF IP addresses in a separate subnet than that of VIPA addresses (IPv4) or using a separate prefix than that of VIPA addresses (IPv6).
- Do not specify static IUTSAMEH links or interfaces.

It is also possible with multiple stacks in the same LPAR to end up with both XCF and HiperSockets links or interfaces. This occurs when the availability of the (preferable) HiperSockets link or interface changes as each TCP stack (within the same LPAR) is started. For example, stack A is started with HiperSockets available and later stack B is started with HiperSockets unavailable. This type of configuration should be avoided.

Efficient routing using HiperSockets Accelerator

Communications Server leverages the technological advances and high performing nature of the I/O processing offered by HiperSockets with the IBM Z® servers and the IBM OSA-Express feature using QDIO architecture by optimizing IP packet forwarding processing that occurs across these two types of links. This function is referred to as HiperSockets Accelerator. It is a configurable option, and activated by configuring the IQDIORouting option on the IPCONFIG statement.

Restrictions:

- HiperSockets Accelerator is IPv4 only.
- You cannot enable HiperSockets Accelerator if you enable IP security on the stack.
- You cannot enable HiperSockets Accelerator if IP forwarding is disabled on the stack.

When configured, it allows unicast IPv4 packets that are received over a HiperSockets link and are to be forwarded over a QDIO link (or received over QDIO and are to be forwarded over HiperSockets) to be forwarded by the z/OS Communications Server HiperSockets device driver. That is, the IP forwarding function is pushed down as close to the hardware [or to the lowest software DLC (Data Link Control)] layer as possible so that these packets do not have to be processed by the TCP/IP stack or address space. Therefore, valuable TCP/IP resources (storage and machine cycles) are not expended for purposes of routing and forwarding packets. Figure 17 on page 85 illustrates a configuration before the use of HiperSockets Accelerator.

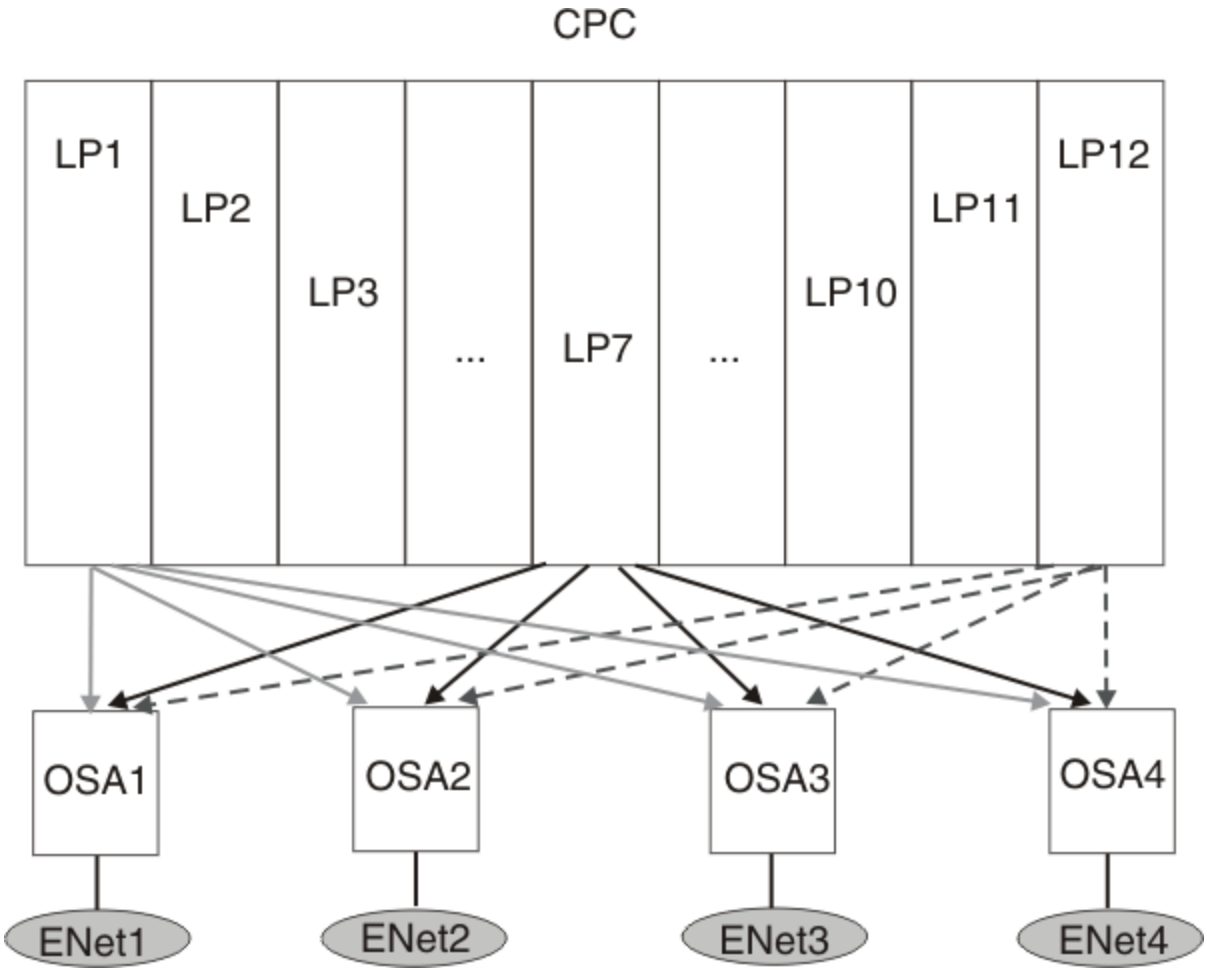


Figure 17. Candidate configuration for HiperSockets Accelerator

HiperSockets Accelerator presents a different configuration and approach to obtain full connectivity, as shown in Figure 18 on page 86.

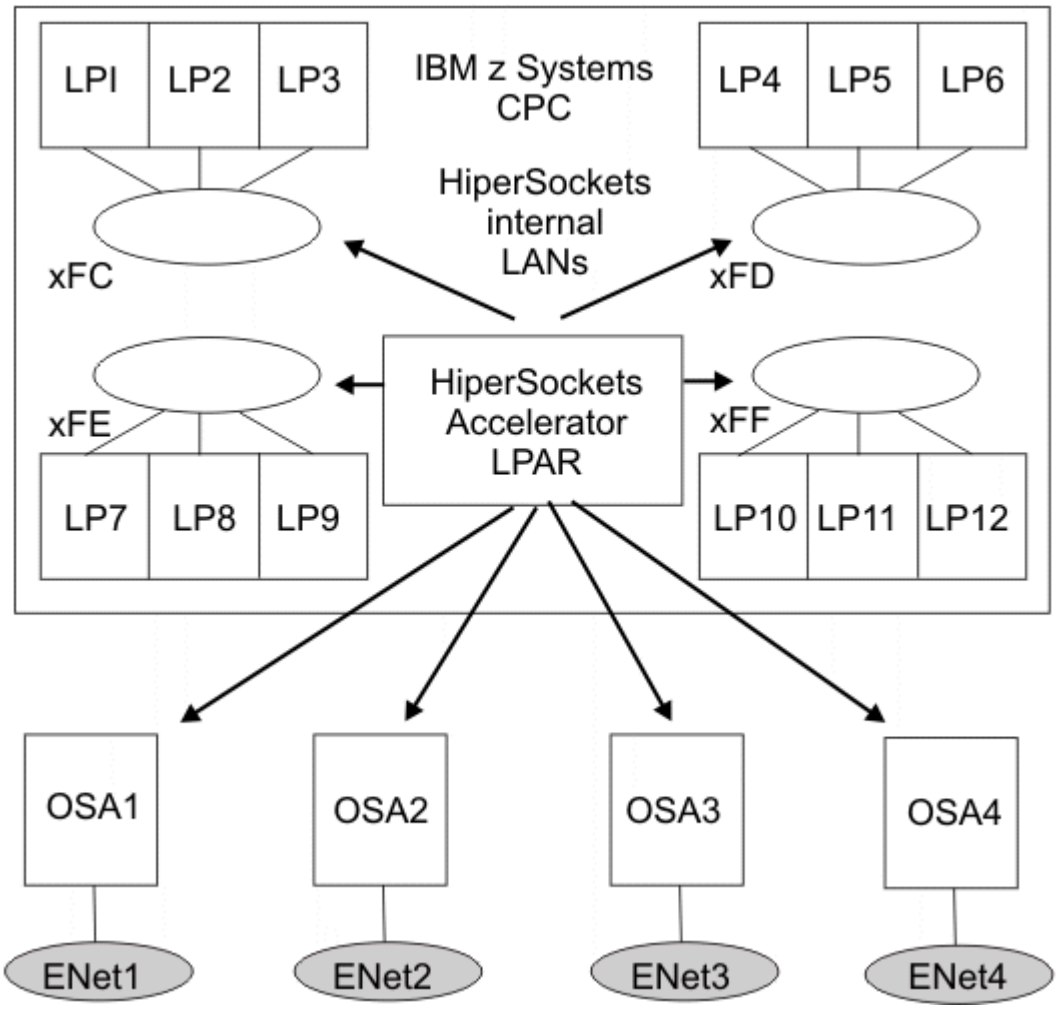


Figure 18. HiperSockets Accelerator configuration

This function allows a user to position a specific or single TCP/IP stack which has direct physical connectivity to the OSAs LANs as the HiperSockets router. This stack can then connect to all remaining TCP/IP stacks in other images (LPARs) within the same CPC that require connectivity to the same OSA LANs using HiperSockets connectivity.

This approach becomes more beneficial as the number of LPARs within a given CPC increase. Instead of attempting to directly attach each LPAR to each physical network attachment using an OSA LAN, a smaller number of OSAs could be concentrated through a single z/OS LPAR. From a performance perspective, HiperSockets Accelerator attempts to make the intermediate (or router) TCP/IP stack appear as if it did not exist in the path. Instead, each LPAR will appear as if each were directly attached to the physical network (for example, packets are forwarded without traversing the router TCP/IP stack). There are no additional routing configuration tasks required by the user. The prerouting occurs automatically. The TCP/IP stack automatically detects IP packet forwarding is occurring across a HiperSockets eligible route (QDIO/HiperSockets or HiperSockets/QDIO), and dynamically creates a HiperSockets Accelerator route entry. All subsequent packets will then take the optimized device driver path, and will not traverse the TCP/IP stack.

The dynamically created HiperSockets Accelerator routing entries can be displayed by the Netstat ROUTE/-r report option with the IQDIO modifier. VTAM tuning statistics are provided to allow the user to monitor or measure prerouting activity.

QDIOPriority (IQDIORouting option) is an optional choice that allows the user to specify which of the four priority queues should be used when prerouting packets from a HiperSockets link outbound to a QDIO link. The default is 1 (highest priority), and in most cases should be sufficient.

For additional details regarding the IQDIORouting configuration option, see the IPCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Tips:

- If packet trace is enabled, then packets that are accelerated by HiperSockets Accelerator do not appear in the packet trace (either inbound or outbound). If you want function similar to that of the packet trace in conjunction with HiperSockets Accelerator, you can use the OSA-Express network traffic analyzer (OSAENTA) to trace these packets that are accelerated to or from OSA-Express QDIO. For more details on OSAENTA, see "OSA-Express network traffic analyzer trace" on page 104.
- QDIO Accelerator provides all of the function supported by HiperSockets Accelerator, and provides accelerated forwarding for other combinations of traffic involving QDIO devices, including sysplex distributor. QDIO Accelerator can also interoperate with IP security. For more information, see "QDIO Accelerator" on page 101.

HiperSockets multiple write

The HiperSockets multiple write facility moves multiple output data buffers in a single write operation. This facility might reduce CPU usage, and might provide a performance improvement for large outbound messages that are typically generated by traditional streaming workloads such as file transfer, and interactive web-based services workloads such as XML or SOAP.

To enable the HiperSockets multiple write facility on all HiperSockets interfaces, including interfaces created for dynamic XCF, add the IQDMULTIWRITE parameter to the GLOBALCONFIG statement. For more information about the GLOBALCONFIG statement and the HiperSockets multiple write facility, see z/OS Communications Server: IP Configuration Reference.

Restriction: HiperSockets multiple write is effective only on an IBM System z10 or later server when z/OS is not running as a guest in a z/VM environment.

HiperSockets multiple write assist with IBM zIIP

When your system is an IBM System z10 or later server and the HiperSockets multiple write facility is enabled, an additional assist for large outbound TCP messages is available through the IBM Z® Integrated Information Processor (zIIP) on a System z9 or later server. To enable HiperSockets write processing for large outbound TCP messages on available zIIPs, specify the ZIIP IQDIOMULTIWRITE parameter on the GLOBALCONFIG statement.

When your system is an IBM System z10 or later server that does not have zIIPs configured, you can use the zIIP HiperSockets multiple write facility to project the percentage of existing HiperSockets workload currently running on central processors that would be eligible to run on zIIPs, if zIIPs were available on the z/OS image. To perform such projection analysis, specify the following parameters on the GLOBALCONFIG statement:

- IQDMULTIWRITE parameter
- ZIIP parameter with the value IQDIOMULTIWRITE

In addition, you must also specify PROJECTCPU= YES in the IEAOPT*xx* member of SYS1.PARMLIB. Then run your HiperSockets workload, and SMF provides accounting information regarding zIIP-eligible workload.

Guideline: Remove GLOBALCONFIG ZIIP IQDIOMULTIWRITE from your TCP/IP profile after you have completed your zIIP performance projection runs. TCP/IP consumes slightly more central processing resources when no zIIPs are online and you have coded GLOBALCONFIG ZIIP IQDIOMULTIWRITE.

For information about configuring the PROJECTCPU parameter in the IEAOPT*xx* member of SYS1.PARMLIB, see z/OS MVS Initialization and Tuning Reference. For information about accounting for zIIP eligibility in SMF record types 30 and 7*x*, see z/OS MVS System Management Facilities (SMF). For information about zIIP-related reporting updates, see z/OS RMF Report Analysis.

z/OS HiperSockets connectivity

"HiperSockets concepts and connectivity" on page 75 describes how in z/OS HiperSockets can be used for two unique purposes with each option that have unique configuration requirements summarized here as follows:

Dynamic XCF (dynamically defined HiperSockets)
Dynamic XCF links provide connectivity to other z/OS instances within the same sysplex. When other z/OS instances within the sysplex are also running on the same central processor complex (CPC), HiperSockets interfaces are (optionally) dynamically created to those instances.

User-defined HiperSockets interface (non XCF-related HiperSockets)
User-defined (static) HiperSockets interface definitions provide connectivity to other z/OS or Linux instances running within the same CPC that are unrelated to sysplex (or the IP subnet of the dynamic XCF interface). In this use case, the administrator manually configures the user-defined HiperSockets IPv4 or IPv6 interface by using DEV/LINK or INTERFACE statements.

HiperSockets uses the Internal Queued Direct I/O (IQD) channel architecture on IBM Z. IQD channel path IDs (CHPIDs) and their related subchannel devices are defined in hardware configuration definition (HCD). The IQD system firmware is based on QDIO system architecture. QDIO architecture defines the following two connectivity modes of internal system connectivity for the internal routing of network traffic as follows:

Layer 3 (L3) mode
IQD firmware routes packets to other hosts using the same IQD CHPID based on the (next hop) IP address.

Layer 2 (L2) mode
IQD firmware routes packets to other hosts using the destination MAC address (unaware of the L3 protocol).

Each IQD CHPID serves as a unique internal (IBM Z) isolated network. Each internal IQD network can optionally be configured with a single physical network identifier (PNetID).

Operating systems running on the same CPC that require HiperSockets connectivity must:

- Connect to the same IQD channel (CHPID).
- Use the same QDIO (L2 or L3) mode.
- Use the same VLAN (when VLANs are used).

z/OS HiperSockets (IQD) interfaces created by both DYNXCF and user-defined interfaces support L3 mode only and the mode is not configurable in z/OS. The next topic introduces a different configuration approach that does use QDIO L2 mode.

HiperSockets Converged Interface overview

z/OS Communications Server provides the capability to integrate HiperSockets connectivity with your external LAN. HiperSockets connectivity converged with the external LAN is referred to as the z/OS Communications Server HiperSockets Converged Interface (HSCI) function.

Within each central processor complex (CPC), you can define a specific Internal Queued Direct I/O (IQD) channel path ID (CHPID) for HiperSockets to provide connectivity that is associated with your designated external LAN. You must configure the designated IQD CHPID by using a channel parameter in the hardware configuration definition (HCD), which enables the IQD External Bridge channel function of HiperSockets. The IQD External Bridge function is a channel function and is not a new CHPID type. When the IQD External Bridge function is configured, the single IQD CHPID is integrated with your designated external LAN; for communications within the CPC, the External Bridge function enables secure and high-performance network access using the designated IQD CHPID.

The z/OS HSCI function transparently converges a HiperSockets interface with your OSA (OSD) interfaces providing transparent and dynamic usage of HiperSockets. HSCI provides z/OS users with a third HiperSockets hands free usage (administrative) model. The HSCI inherits all of the OSA network-related attributes (for example, IP route, IP address, subnet, VLAN, VMAC, and so on).

The HSCI configuration option is managed with a single TCP/IP global configuration option called AUTOIQDC (automatically create, converge, operate, and manage the IQD interface). HSCI is intended to replace the (need for) static user-defined HiperSockets interfaces. The HSCI function is unrelated to and independent of DYNXCF-related HiperSockets.

The key benefits of the HSCI are:

- Combines the existing high-performance attributes of HiperSockets for intra-CPC communications with your external LAN providing a single IP network administrative model.
- z/OS HiperSockets administrative usability improvements provide:
  - Dynamic configuration and reconfiguration of HiperSockets interfaces.
  - Automatic provisioning and connectivity to the IQD network without requiring operational actions such as Start or Stop Interface commands.
  - Dynamic discovery of and communications with peer hosts that have access to the same IQD network.
- Converges OSA connectivity and HiperSockets connectivity into a single logical network interface. The single interface simplifies network management by reducing the number of resources that are required to represent individual unique networks, including the following resources:
  - IP interfaces
  - IP subnetworks (VLANs when applicable)
  - IP addresses
  - IP routes
- z/OS mobility improvements:

  When a z/OS instance is moved to another CPC as your OSA interfaces are started, TCP/IP automatically and dynamically reconnects to HiperSockets on the new CPC without requiring any TCP/IP configuration or configuration changes.
- Linux Layer 2 (L2) mode connectivity for HiperSockets providing compatibility with Linux instances using the z/VM Virtual Switch Bridge environment (requires L2 mode).
- Enables sysplex network traffic to transparently use HiperSockets connectivity for VIPAROUTE processing over OSD interfaces.
- Single IP network topology:

  HSCI provides a single seamless LAN topology where your internal IQD networks within each CPC are no longer unique and isolated IP subnets. With the HSCI function, your IQD network becomes part (an extension) of your external LAN (creating a single IP subnet). This architectural model is similar to what a switch Bridge port provides, bridging separate networks and creating a single logical network topology. The single network model simplifies many network administrative tasks. The concepts of the single network are illustrated in the following figures.

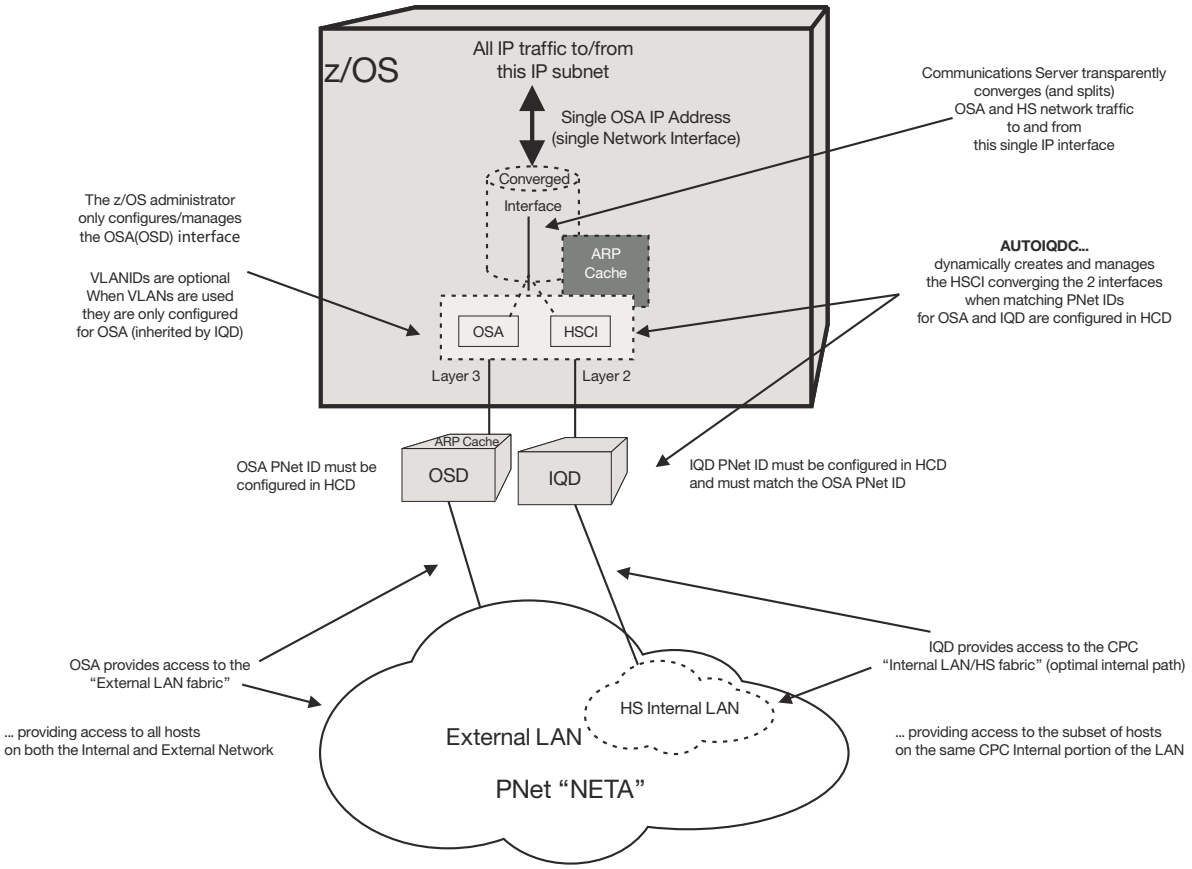


Figure 19. HiperSockets Converged Interface overview

Figure 19 on page 90 shows the key concepts of the HSCI function. HSCI logically converges the network interfaces provided by IQD with OSD CHPID types into a single logical interface. HSCI is transparent to the application layer. The upper layers of the TCP/IP software stack only have visibility to a single OSA interface that has a single IP address and IP subnet.

The z/OS Communications Server HSCI function is controlled with a single TCP/IP Global Configuration parameter called AUTOIQDC. The AUTOIQDC parameter indicates that Communications Server should automatically create, manage, and converge the HSCI with the associated OSD interfaces. Communications Server associates the dynamically generated IQD interface with an OSA (OSD) interface based on the IQD and OSD CHPIDs that have matching PNetIDs.

Other than the AUTOIQDC TCP/IP parameter there are no other TCP/IP or VTAM definitions that are required to use the HSCI function. However, the associated OSA (OSD) configuration must be defined with the following:

- The INTERFACE statement (cannot be defined with DEVICE/LINK).
  - An IPv6 INTERFACE cannot share an OSA port with an IPv4 DEVICE/LINK.
- VMAC (OSA generated).

The AUTOIQDC setting enables the dynamic creation of the HSCI and the VTAM TRLE when an OSA interface is started. The IQD CHPID I/O device definitions are required in HCD. For each unique IQD CHPID Communications Server creates a single TRLE for IPv4 and a separate (single) TRLE for IPv6. Each HSCI TRLE uses up to 10 devices (eight data devices and two control devices).

Communications Server dynamically determines whether traffic routed to the external LAN over a given OSD CHPID can also use the HSCI; if the traffic can use the HSCI, Communications Server routes the traffic over the high-performance HiperSockets IQD CHPID. If the AUTOIQDC parameter on the GLOBALCONFIG statement in the TCP/IP profile is specified, when the first OSD INTERFACE is started the HSCI, also known as the IQDC interface, and TRLEs are dynamically created.

The naming convention for the dynamically created:

- HSCI is EZAIQCxx (IPv4) and EZ6IQCxx (IPv6), where xx is the IQD CHPID number to which a particular HSCI is associated. The HSCIs are identified by interface type IPAQIQDC for IPv4 and interface type IPAQIQDC6 for IPv6. Traffic for unique VLANs over the same OSD CHPID (multiple VLANs) share the same HSCI.
- HSCI TRLEs is IUTIQCxx (IPv4) and IUTIC6xx (IPv6), where xx is the IQD CHPID number to which a particular HSCI is associated.

Routing, IP security, and packet tracing of the dynamic HSCI is controlled through their associated OSD interfaces; the HSCI is generally not visible to these functions.

- Routing daemons such as OMPROUTE are aware of only the associated OSD interfaces; the stack does not report IQDC interfaces to OMPROUTE, and static routes cannot reference the dynamic IQDC interfaces. z/OS Communications Server decides to direct traffic over an IQDC interface after the associated OSD interface is chosen as the route for the traffic.
- HSCI is not directly visible to IP security. If IP security is enabled, traffic over an HSCI inherits the security class (SECCLASS) from the associated OSD interface. For packets that are sent and received over the HSCI, IPSec filtering applies to the security class of the associated OSD interface.
- Packet tracing of an OSD interface includes traffic that is redirected or received over the HSCI. If you perform a packet trace for data over an HSCI, all outbound packets are traced as if they were sent over OSD. Inbound packet trace has a flag that indicates the packet was received over the HSCI.

You can use Netstat reports to display information about dynamically created IQDC interfaces.

- The Netstat DEvlinks/-d report displays information for each IQDC interface, as it does for any defined interface. The Netstat DEvlinks/-d report for a given OSD interface also displays how much traffic was sent and received over the associated IQDC interface. For more information about the Netstat DEvlinks/-d report, see z/OS Communications Server: IP System Administrator's Commands.
- You can monitor which IP addresses are reachable over the OSD interfaces and over the associated HSCI by using the Netstat ARp/-R and Netstat ND/-n reports. For more information about the Netstat ARp/-R report and the Netstat ND/-n report, see z/OS Communications Server: IP System Administrator's Commands.
- You can display information about the dynamic HSCI TRLEs and their associated IQD datapath devices by issuing the **D NET,ID=trle** command or the **D NET,TRL,TRLE=trle** command. For more information about the DISPLAY ID command and the DISPLAY TRL command, see z/OS Communications Server: SNA Operation.

Using the hardware configuration definition (HCD) you can define up to 32 unique IQD channel path IDs (CHPIDs) within each CPC. Each IQD CHPID within a CPC functions as a unique internal isolated network. Each IQD CHPID (network) can optionally support a single unique PNetID. HSCI requires an IQD PNetID to be configured on the IQD CHPID.

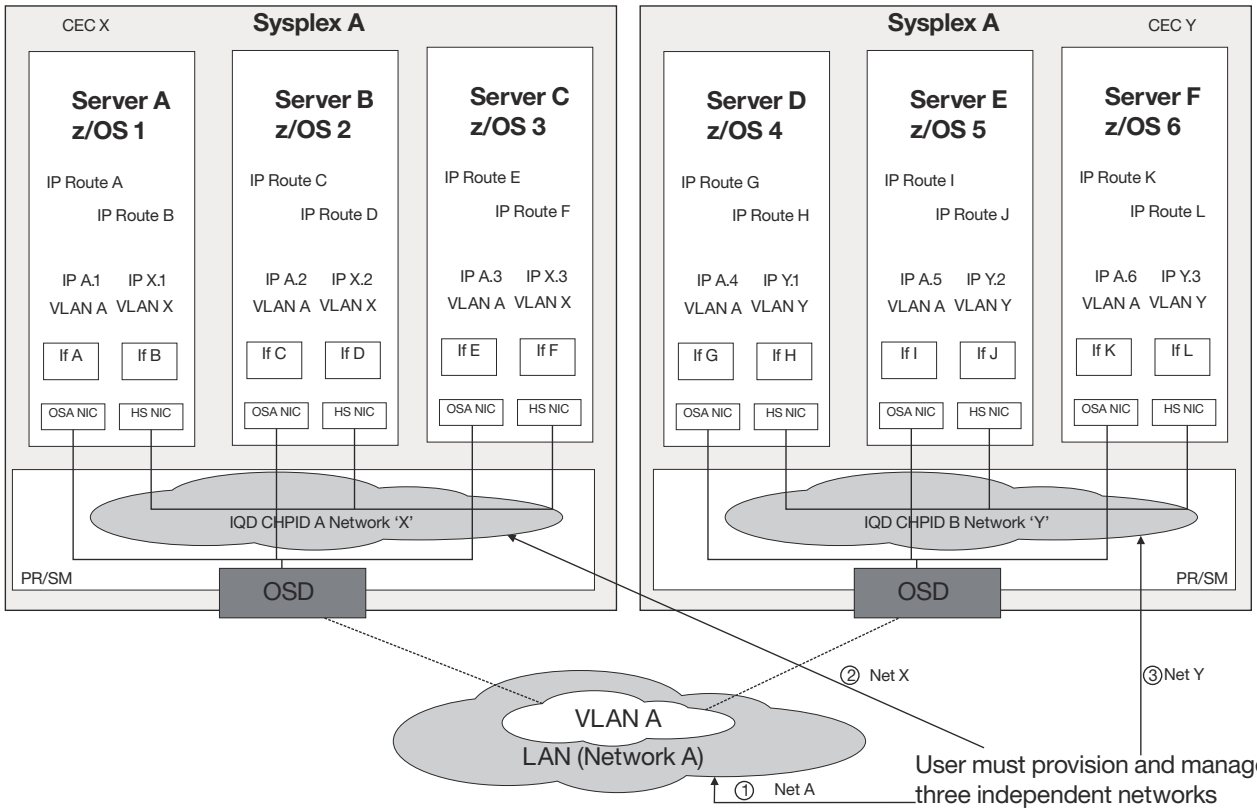


Figure 20. Managing HiperSockets networks without HSCI

Figure 20 on page 92 illustrates two CPCs (CECs X and Y) each with a single IQD network. Both systems also have access to the same external LAN (network A). In this basic (non-sysplex) network example, the administrator must manage three independent networks (Networks X, Y, and A). Networks X, Y, and A also represent (the concepts of) three unique physical L2 networks each having a unique PNetID. Here the term physical means each network (broadcast domain) is physically (logically) isolated from the other two networks. Each physical network could optionally be subdivided into multiple separate L2 networks by defining (multiple) IP subnets each with corresponding VLAN IDs (for example, as in VLAN A). Connecting hosts across multiple unique L2 networks requires a Layer 3 (L3) IP router.

In this configuration, each z/OS instance is shown to have two IP interfaces each having unique IP addresses, VLAN IDs and IP subnets (subnet masks). Each corresponding IP interface also has a unique (static or dynamic) IP route that is defined for directing traffic over each associated interface. The network administrator creates three unique networks requiring that each network and all associated network resources must be independently provisioned and managed. As the number of IP network interfaces increases, the complexity of network administration also increases.

In this example, if an instance of z/OS is to be moved (for example, z/OS 3 moved from CPC X to CPC Y), the HiperSockets interface needs to be reconfigured to access Network Y (in CPC Y) instead of Network X (in CPC X).

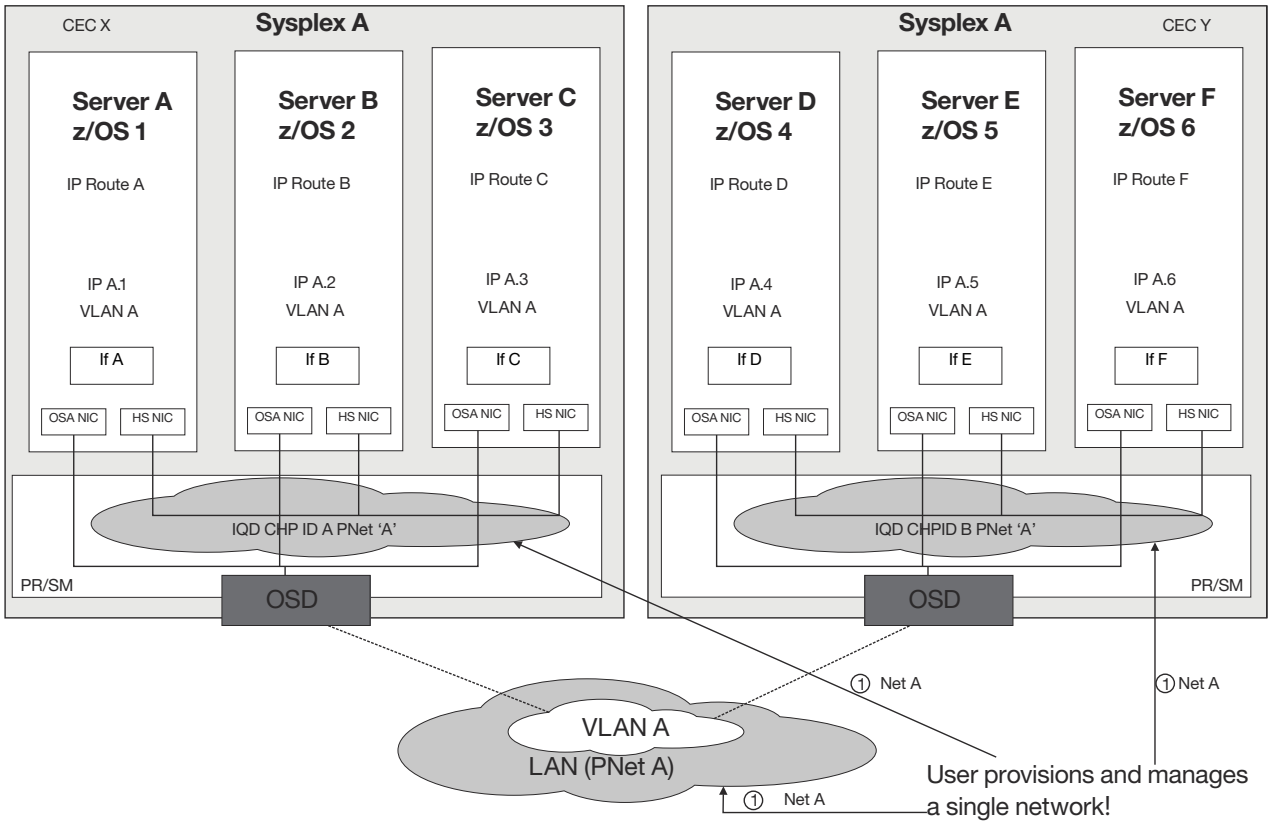


Figure 21. Managing HiperSockets networks with HSCI

When using the HSCI function the HiperSockets interfaces are no longer visible to, configured, or managed by the z/OS administrator. When z/OS is using the HSCI function, it can only communicate with other z/OS instances over HiperSockets that are also using the HSCI function.

When contrasting Figure 20 on page 92 with Figure 21 on page 93, many of the key advantages of HSCI are readily seen and summarized as follows:

- HSCI creates a single internal and external IP network.

  Instead of provisioning and managing three unique networks, two internal and a third external network for the two systems, the administrator only provisions and manages a single IP network (network A).
- HSCI eliminates all resources that are related to the HiperSockets networks.

  Each z/OS instance has access to a single IP network (network A), defined by the OSA interface thereby eliminating the HiperSockets interface related resources, such as the HiperSockets, IP route, IP address, IP interface, IP subnet, and VLAN.
- z/OS mobility is streamlined.

  Therefore, when moving an instance of z/OS (for example, z/OS 3 from CPC X to CPC Y), the HiperSockets interface (HSCI) is not visible to z/OS, HiperSockets no longer must be reconfigured. HSCI automatically and dynamically provides access to HiperSockets on the new CPC.
- The HSCI function uses HiperSockets in L2 mode only.
- Using HSCI with multiple (unique) external LANs.

  The z/OS HSCI function supports integration with multiple external LANs. Figure 21 on page 93 illustrates a use case that has a single external LAN (Net A). In cases where a customer has multiple unique external LANs (for example, multiple LANs that require isolation, such as Net A, Net B, Net C) and you want to use the HSCI function for more than one LAN or each unique external LAN, you can provision a unique IQD CHPID for each unique external LAN. In this example, you can provision another IQD CHPID for each unique LAN defining each IQD CHPID with the corresponding PNetID (Net A, Net

B, Net C). The OSD interfaces that are associated with each unique external LAN will require matching PNetIDs to correspond with the IQD PNetIDs. IBM Z supports up to 32 unique IQD CPHIDs per CPC.

- z/OS to z/OS HSCI migration considerations.

  When z/OS is using the HSCI function, it can only communicate with other z/OS instances over this same IQD CHPID that are also using the HSCI function. If your z/OS system must communicate with down-level z/OS instances over HiperSockets, you are required to use the L3 HiperSockets support that is provided by:

  – Non-sysplex: user-defined HiperSockets interface statements (for z/OS instances that are not within the same sysplex).
  – Sysplex: HiperSockets interfaces that are created by DYNXCF (for z/OS instances within the same sysplex).

HSCI in other configurations

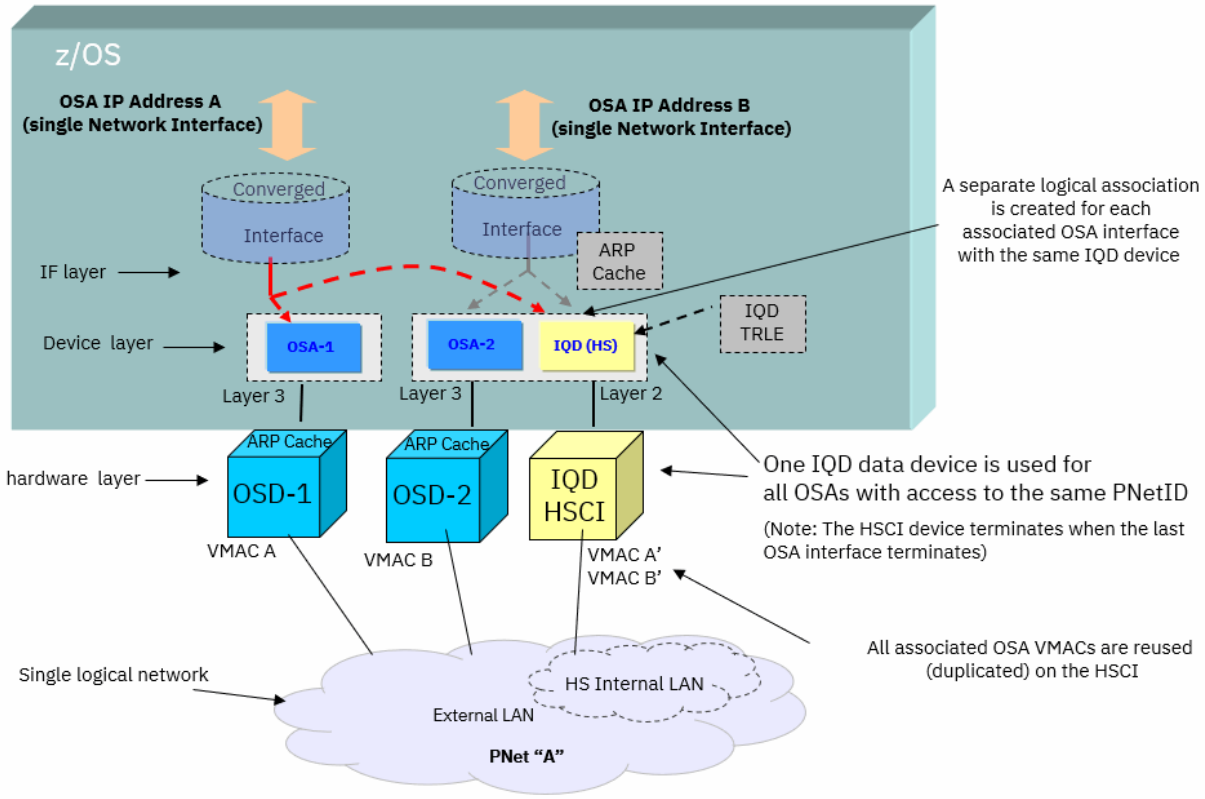


Figure 22. An HSCI sample configuration with two OSA-Express features, two IP interfaces and a single subnet

Figure 22 on page 94 illustrates that when two logical IP (OSD) interfaces share the same PNetID, they will converge with the same HiperSockets (IQDC) data device. No matter how many OSA interfaces share the PNetID, the same HiperSockets data device is used (i.e, the same IQDC interface and the same TRLE are used for all the OSA interfaces with the same PNetID). The first associated OSA activates the IQDC interface and the last OSA that terminates will terminate the IQDC interface.

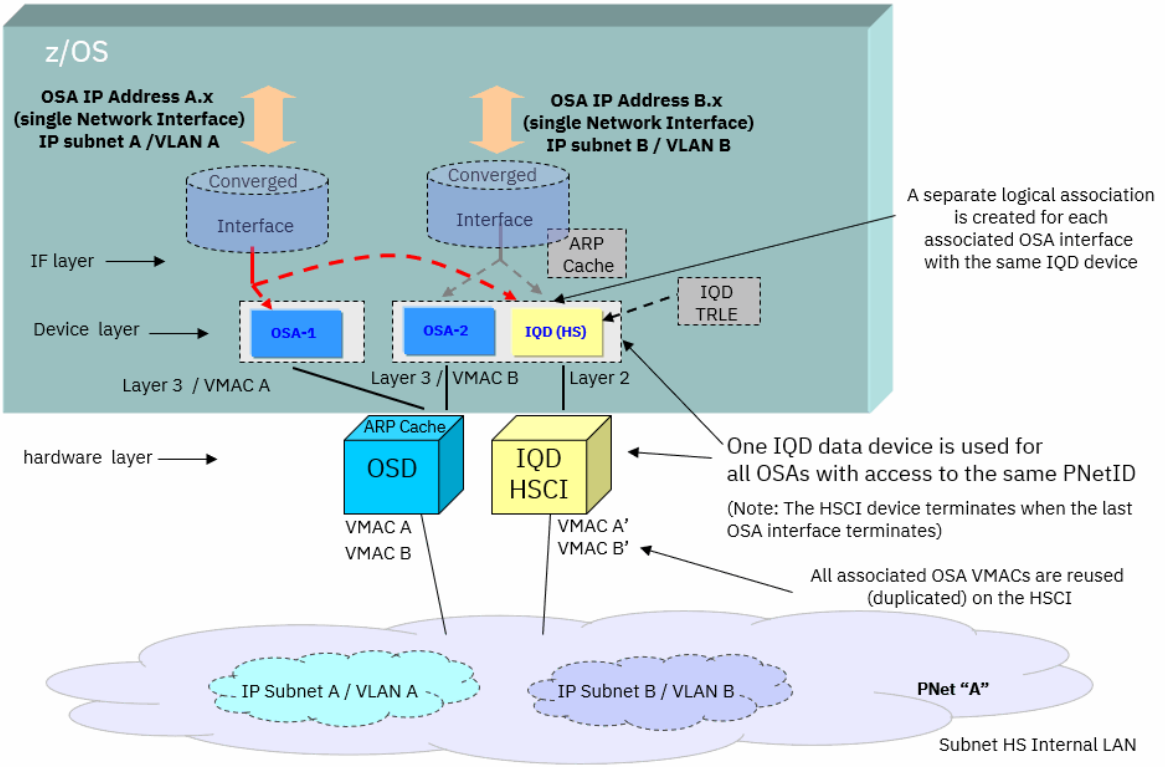


Figure 23. An HSCI sample configuration with a single OSA-Express feature, two IP interfaces and two subnets

Figure 23 on page 95 illustrates that when two logical IP (OSD) interfaces are configured to the same physical OSA-Express feature, they share the same PNetID and converge to the same IQD data device. The IP interfaces in this example are defined with unique VLANs and subnets. No matter how many OSA interfaces are defined for the same physical OSA-Express feature, the same HiperSockets (IQDC) data device is used (i.e, the same IQDC interface and the same TRLE are used for all the OSA interfaces). The first associated OSA interface activates the IQDC interface and the last OSA interface that terminates will terminate the IQDC interface.

Linux and z/VM VSwitch bridge considerations

In addition to the administrative advantages provided by the HiperSockets Converged Interface (HSCI) function for the z/OS environment, the z/OS HSCI function also provides unique advantages for HiperSockets connectivity when connecting to Linux in the z/VM VSwitch bridge environment.

Linux on IBM Z administrators prefer the administrative and operational advantages that are offered by Linux configured with a single IP interface using a (QDIO) Layer 2 (L2) mode. Conversely, managing multi-home (multiple IP interfaces) Linux instances can increase operational complexity. This complexity increases, as the number of Linux guests increase. This (single IP interface) objective is preserved by the z/VM VSwitch bridge support.

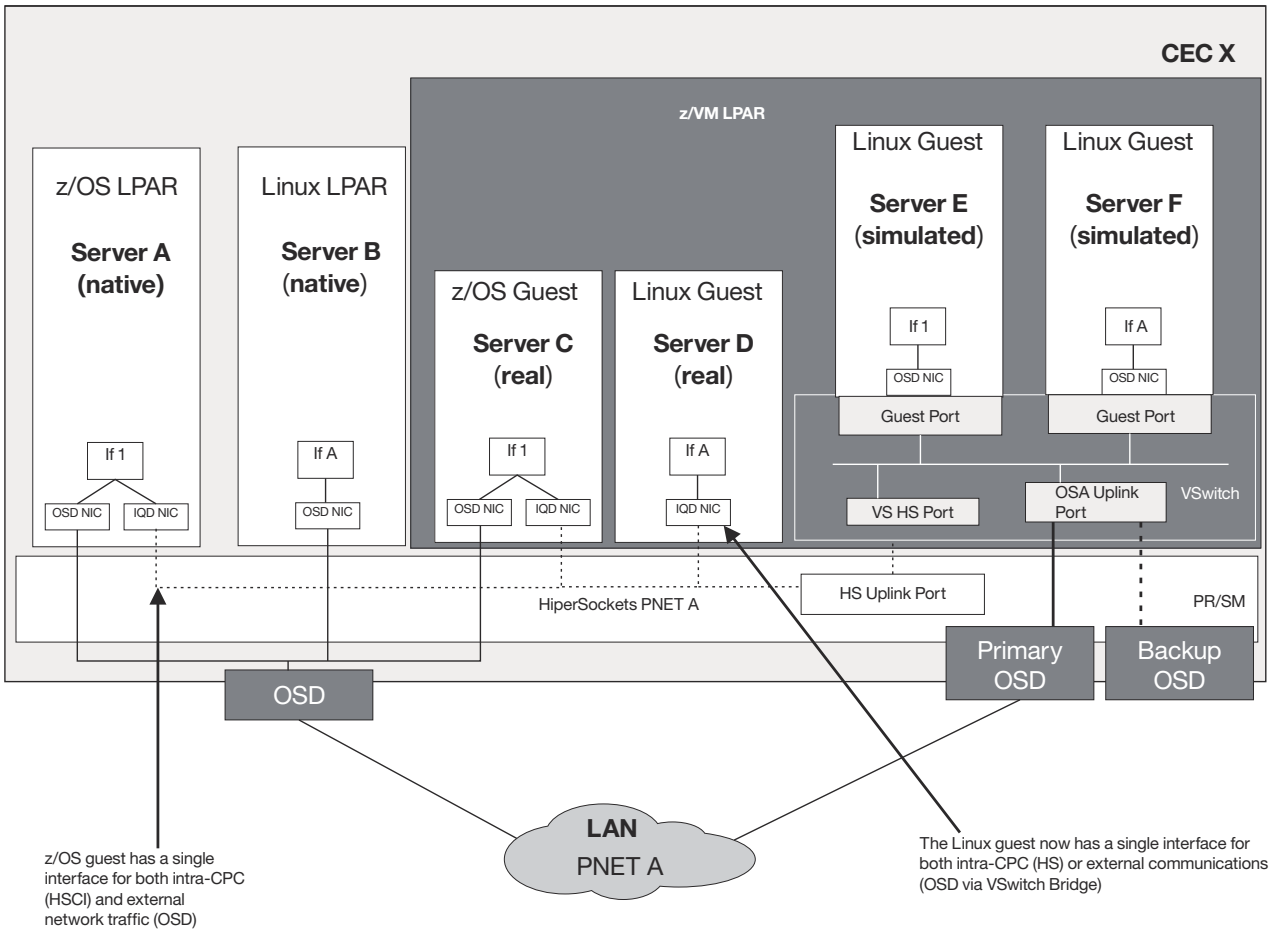


Figure 24. Linux and z/VM VSwitch bridge environment

Figure 24 on page 96 provides an example of the Linux and z/VM VSwitch bridge environment. It illustrates how the VSwitch bridge function allows Linux guests (Server D) to be configured with a single HiperSockets interface that provides both intra-CPC communications (HiperSockets to Servers A and C) and external communications via the VSwitch bridge port through OSA. The Linux usability advantages of the single IP interface are extended by the allowing Linux guests to configure a single IP interface for HiperSockets providing both internal central processor complex (CPC) and external LAN communications.

Starting with IBM z14, Linux native LPARs can also (transparently) use the z/VM VSwitch bridge support. Any z14 operating system (running as a z/VM guest or native LPAR) that connects to an Internal Queued Direct I/O (IQD) channel configured with the External Bridge option (in HCD) is automatically bridge eligible by default. z/OS connects to an External Bridge IQD channel only when configured with AUTOIQDC (HSCI). With AUTOIQDC, z/OS uses OSA for external connectivity and the HSCI for internal connectivity. This provides bridge compatibility to communicate with other operating systems on the same IQD channel, but z/OS always "opts out" of the external bridge capability.

Note the following important network configuration concepts that are shown in Figure 24 on page 96:

- The IQD channel path ID (CHPID) and the OSD PCHPID (port) were configured with the same physical network ID (PNetID = NETA), both providing access to the same physical network.
- The IP topology for both interfaces (in every host) is also configured such that both the HiperSockets and the OSA interfaces are attached to the same IP subnet (and VLAN ID when applicable).
- From the z/OS (servers A and C) HSCI perspective, every host that is reachable via the internal HSCI must also be reachable via OSA (packets are transmitted over both interfaces to the same peer hosts, for example to Server D).

  This means when connecting to z/OS using the HSCI function to Linux, that Linux instance (guest or native LPAR) must be configured to use the same IQD CHPID with the VSwitch bridge support.

- In z/OS if the HSCI terminates, OSA can still be used to communicate with peer hosts. When the last OSA interface terminates, network connectivity is lost (the HSCI is not usable without the OSA interface).
- The HSCI supports (QDIO) L2 mode only. When the IQD system firmware is operating in L2 mode, the firmware is unaware of the Layer 3 (L3) protocol (IP protocol or version).
- z/OS and z/VM VSwitch bridge compatibility:
  - z/OS HSCI is compatible with the z/VM VSwitch bridge environment (z/OS can communicate with Linux guests or Linux LPARs over HiperSockets that are exploiting the bridge, such as Server D).
  - While z/OS with HSCI is compatible with the bridge environment, z/OS is not bridge capable, which means z/OS does not use the z/VM VSwitch bridge (z/OS network traffic does not traverse the z/VM VSwitch bridge port). With the HSCI function, external network connectivity is provided by the converged OSD instead of the bridge function.

    For example, in Figure 24 on page 96, z/OS servers A and C can communicate with servers E and F via external connectivity via OSA only (HiperSockets cannot be used because it does require z/OS being bridge capable).
  - Conversely, the z/OS HiperSockets support that is not related to HSCI only supports L3 mode connectivity which is incompatible with Linux in the z/VM VSwitch bridge environment. The z/OS HSCI support resolves this incompatibility by providing compatibility for HS L2 mode for Linux guests or LPARs using HS with the z/VM VSwitch bridge.
  - Linux supports IQD connectivity in both L2 and L3 modes. The z/VM VSwitch bridge requires (supports) L2 mode only.

  For more information about the z/VM VSwitch bridge support, see *z/VM Planning and Administration*.
- z/OS to Linux HSCI migration considerations:

  When z/OS uses the HSCI function, it can only communicate with Linux guests or LPARs using HiperSockets configured to use the z/VM VSwitch bridge. If your z/OS system is to communicate with Linux instances that are not configured to use the z/VM VSwitch bridge, you are required to use the L3 HiperSockets support that is provided by z/OS user-defined HiperSockets INTERFACE statement.

Performance considerations for HiperSockets Converged Interface

For network traffic within a central processor complex (CPC), HiperSockets can provide lower latency when compared to external networks. Lower latency is most likely beneficial for interactive (request and response) workloads where latency (response time) is critical.

However, when large amounts of data are transmitted (typically, streaming or bulk data workloads) as a significant portion of the intra-CPC workload, you might prefer OSA-Express processing because of differences in the OSA-Express processing for large data transmissions (for example, segmentation offload). The savings that are offered by OSA-Express processing for large data transfers vary. HiperSockets transmissions generally use more processor cycles than OSA transmissions, and this difference in processing cycles is magnified as transmission size grows. If you prefer using OSA-Express processing for large transmissions, while still using HSCI, then this variation is also supported.

The HiperSockets Converged Interface (HSCI) function uses the strengths of both the HiperSockets and OSA-Express technologies, by providing a configuration option that transparently controls traffic selection. Specifying NOLARGEDATA on the AUTOIQDC parameter of the GLOBALCONFIG statement transparently directs large data transmissions over OSA-Express for HSCI while directing small data transmissions over Internal Queued Direct I/O Converged (IQDC) interfaces. You can enable this setting to get the lower latency of HiperSockets for small data transmissions, while avoiding the processor cost penalty sometimes incurred with HiperSockets for large data transmissions. If you do not anticipate a significant number of large data transmissions or you always prefer the internal processing of HiperSockets (regardless of processor cycle consumption), you should use the default setting (do not specify NOLARGEDATA on the AUTOIQDC parameter).

Tips:

- IQDMULTIWRITE provides optimization when sending large transmissions or a high volume of traffic over HiperSockets. When IQDMULTIWRITE is enabled on the GLOBALCONFIG statement, the IQDC interface also uses this feature. You can optionally disable IQDMULTIWRITE on the GLOBALCONFIG statement (which disables multiple write for all HiperSockets devices).
- The QDIO Accelerator function provides accelerated forwarding at the DLC layer for inbound packets over HiperSockets (IQDC interfaces) that are forwarded outbound over OSA-Express QDIO. For more information, see "QDIO Accelerator" on page 101.

Guidelines:

- When NOLARGEDATA is specified for the AUTOIQDC parameter, Communications Server forces large TCP transmissions to the OSD path. Communications Server defines a large transmission as a TCP socket send of 32 K or larger. All other eligible traffic is transmitted over IQDC interfaces.
- When an OSA adapter is shared among two or more logical partitions (LPARs), the processor savings offered by the OSA adapter are less than the savings offered by configurations that use multiple OSA adapters that are not shared. For example, some OSA assist functions, such as segmentation offload, are not available for the shared OSA configuration.
- When workloads which transmit large data (socket applications issue socket send of 32 K or larger) are a predominant part of the overall network workload and you use HS, you should consider using:
  - The IBM Z Integrated Information Processor (zIIP) function for HiperSockets, which minimizes the general-purpose processor usage that is incurred with HiperSockets.
  - A configured Internal Queued Direct I/O (IQD) maximum frame size of 64 K in hardware configuration definition (HCD).
- When transmitting packets, the HSCI uses the OSA MTU size. Jumbo frames are 9 K (8992) and can be used by TCP connection over the OSA interface. Jumbo frames are too large for the default IQD frame size (16 K supporting an 8 K MTU). To be sure to take full advantage of IQDC for streaming workloads, consider using an IQDC frame size larger than 16 K if you use an MTU size larger than 8192 bytes for OSD. Otherwise, certain streaming workloads are not able to use the HSCI.

SMC and HSCI PNetID considerations

Shared Memory Communications (SMC) and the HiperSockets Converged Interface (HSCI) function are independent functions that can be used independently of or concurrently with each other. The two functions have some subtle overlap that must be considered when configuring the physical network IDs (PNetIDs).

When reviewing the following PNetID concepts and rules, see examples in the previous figures.

The following list provides a summary of the PNetID background concepts and usage rules:

- PNetIDs represent a user-defined logical ID (name) of a physical network. The physical network represents the physical broadcast domain of a given network.
- PNetIDs are defined on a physical port of a physical adapter indicating the ID of the network that this port is associated with (connected to).
- PNetIDs are used by z/OS to associate related adapters with each other. When physical ports have matching PNetIDs, it indicates that the adapters have connectivity to the same physical network and the IDs are used by z/OS as follows:
  - Shared Memory Communications over Remote Direct Memory Access (SMC-R): Associate OSA ports with RoCE ports.
  - Shared Memory Communication - Direct Memory Access (SMC-D): Associate Internal Queued Direct I/O (IQD) channel path IDs (CHPIDs) with Internal Shared Memory (ISM) CHPIDs or OSA ports.

    Note: IQD and ISM CHPIDs are virtual adapters (provided by system firmware) that do not have physical ports. Logically (defined by system architecture), each virtual adapter (CHPID) has a single logical port (supporting a single PNetID per CHPID).
- PNetIDs rules without the HSCI function.

Without the HSCI function, OSA (OSD) and IQD networks cannot have matching PNetIDs since the interfaces to those networks connect to separate and unique isolated networks.

Note: The definition of "without HSCI" means Layer 3 (L3) HiperSockets interfaces that are created by both of the following:

- User-defined HiperSockets interfaces.
- Dynamically defined HiperSockets interfaces that are created by DYNXCF.

The HSCI function logically combines IQD and OSD networks into a single logical network (see examples in Figure 21 on page 93 and Figure 24 on page 96). The following list describes how the HSCI function impacts the configuration and usage rules of PNetIDs:

- PNetIDs with the HSCI function:
  - The HSCI function is applicable when the following conditions are true:
    - HSCI is enabled for this TCP/IP stack (global configuration).
    - An IQD CHPID is available to this z/OS instance, and the CHPID is configured in HCD with both of the following:
      - The External Bridge parameter.
      - A PNetID that matches an OSA PNetID.
  - When the HSCI function is not used (not applicable):

    ISM can be associated (have a matching PNetID) with either IQD or OSD, but not both (IQD and OSA are unique networks and cannot have the same PNetIDs).
  - When the HSCI function is used (applicable):

    ISM can be associated (have a matching PNetID) with both IQD and OSD (IQD and OSD are now converged into a single physical network, supporting a single PNetID).

    With the HSCI function, OSD, RoCE, IQD, and ISM can all have the same PNetID.

    Note: The z/VM VSwitch bridge support also combines (bridges) IQD to OSD networks together forming a single logical network. In the bridge configuration, the PNetIDs can also be equal (forming a single network). z/OS does not directly use the z/VM bridge function (z/OS is not bridge eligible). z/OS is compatible with the bridge function when using the HSCI function. Without the HSCI function, z/OS is not aware of or does not validate PNetIDs for bridge configurations.
- Considerations for dynamic changes in system I/O configuration:
  - HCD definitions (PNetID values and External Bridge parameter) can be changed using dynamic I/O CHPID operations. Changing the IQD CHPID in HCD can alter the PNetID rules (enabling or disabling the HSCI function).
  - z/OS evaluates the PNetID values and applies the PNetID rules when activating an applicable network interface (interfaces that support PNetIDs).
  - During the interface activation process, z/OS applies the PNetID usage rules based on the current state and applicability of the HSCI function.
  - Once an interface is activated, changing the state of the HSCI function has no effect on an already active interface. You must stop an active interface and then restart the interface to pick up any changes in the HSCI function.

Steps for enabling HiperSockets Converged Interface

It is recommended that users begin by reviewing the HiperSockets Converged Interface (HSCI) background information that is provided in this section along with your anticipated IP connectivity and topology requirements. Your use case for the HSCI function can be for a z/OS only environment or a z/OS environment that also includes Linux using the z/VM VSwitch bridge function.

When z/OS is using the HSCI function, the dynamically created HiperSockets interface operates in Layer 2 (L2) mode only. z/OS HSCI can only connect to other z/OS instances over the HSCI when the peer

z/OS instances are also using HSCI. An Internal Queued Direct I/O (IQD) channel path ID (CHPID) that is configured in hardware configuration definition (HCD) with the external bridge parameter cannot connect to z/OS using Layer 3 (L3) (z/OS user-defined or DYNXCF) HiperSockets.

It is helpful to create a design diagram of your new or changed IP network that also includes a view of your external LAN. Your network design can be used to determine and validate if the z/OS HSCI function can be used in your network environment. When you conclude that the z/OS HSCI function should be used, you must start by completing the required hardware configuration definition (HCD) configuration steps for your Internal Queued Direct I/O (IQD) channel path ID (CHPID). For the complete HCD configuration details, see z/OS HCD User's Guide. The following overview information is a guide to bridge into the related HCD topic.

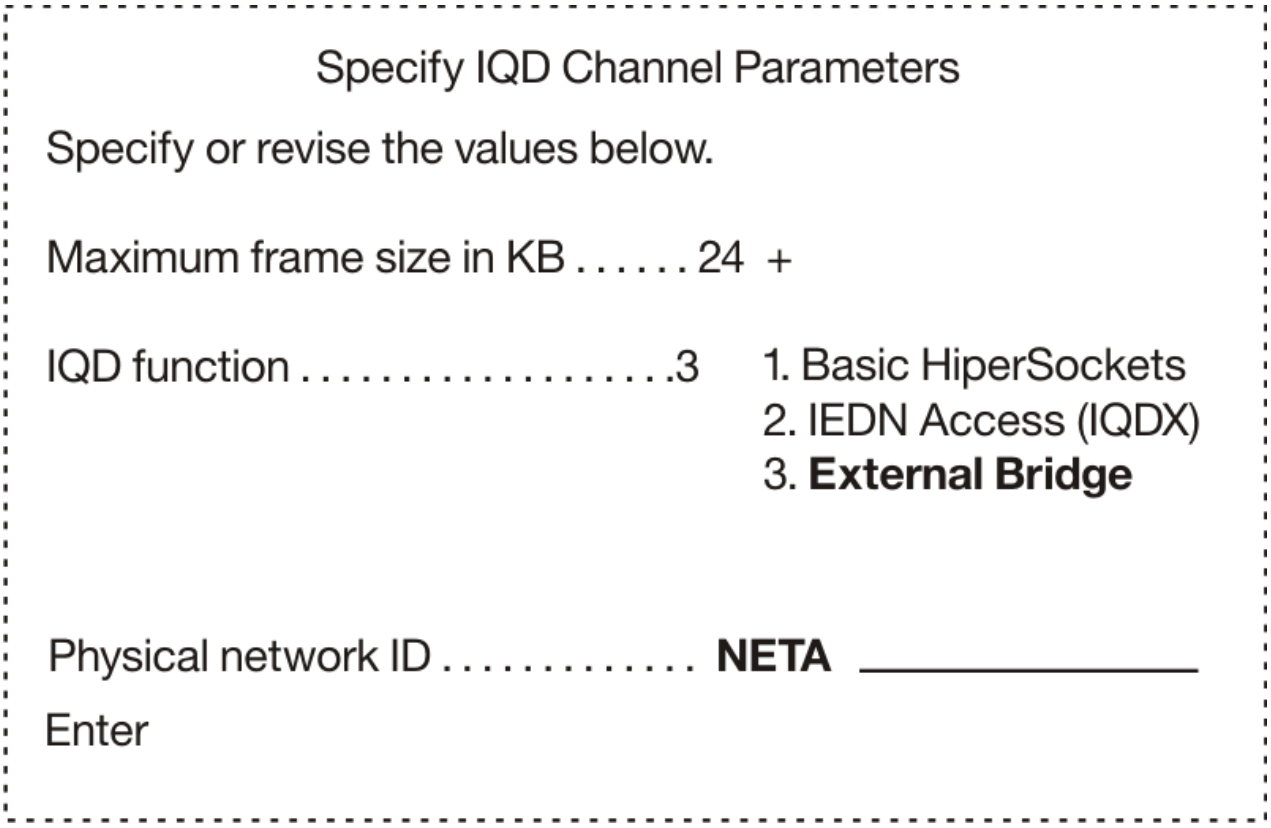


Figure 25. HCD IQD configuration panel (channel parameters)

Tip: To avoid fragmentation issues with jumbo frames, use a maximum frame size greater than 16.

IBM Z hardware configuration definition procedure

1. Determine which IQD CHPID that you plan to use for HSCI. The selected IQD CHPID must be coordinated with all planned users (z/OS and Linux instances) of this IQD CHPID.
2. Complete the HCD configuration steps to define the necessary CHPID and devices that are required for all applicable LPARs. You need to determine the number of LPARs (z/OS and if applicable Linux guest in the z/VM LPARs) that require connectivity over this same IQD CHPID.

   z/OS dynamically finds the IQD CHPID and devices and then creates the HSCI TRLE allocating as many IQD devices for the TRLE as required.

   Note: Communications Server creates an HSCI TRLE for IPv4 and IPv6 (when applicable). Per IP version (IPv4 and IPv6) the z/OS HSCI TRLE requires a minimum of three subchannel devices. When enough devices are configured in HCD, z/OS uses up to a maximum of 10 devices per TRLE (two control and eight data devices). Each z/OS TCP/IP stack (up to eight) uses a single IQD device for all HSCI connectivity (per IP version). The single HSCI IQD device supports all associated OSA interfaces

for the HSCI PNetID (all interfaces that are created for unique physical OSAs and interfaces that are created for multiple VLANs over the same OSA).

3. In HCD define the IQD CHPID to support both of the following:
   - External Bridge

     Note: The External Bridge setting is a required setting for HSCI. This is a system-wide setting informing all users of this IQD CHPID that special rules apply (this IQD CHPID is L2 only).

     This setting does not indicate (require) that the z/VM VSwitch bridge is used.
   - Define the IQD PNetID to match your OSAs PNetID.

     Note: The OSA (OSD) PNetIDs must also be configured in HCD for each associated OSA port.

z/OS Communications Server procedure

When the HCD configuration steps are complete for both IQD and OSD CHPIDs, you can enable the HSCI function in z/OS.

1. In your TCP/IP profile GLOBALCONFIG statement, specify AUTOIQDC. Determine whether you want large data to use your HSCI. See "Performance considerations for HiperSockets Converged Interface" on page 97.
2. Start your OSA interfaces that are associated with your IQD CHPID (OSDs having the same PNetID as the IQD CHPID).

   Use Netstat to verify that the HSCI is automatically activated with your OSA interface. As the HSCI is activated in other z/OS instances using this same IQD CHPID use Netstat to verify that your eligible network traffic is using the HSCI.

   The first associated OSA interface that activates will dynamically create the HSCI TRLE and activate the HSCI. As other associated OSA interfaces are activated, the existing HSCI is discovered and used. The last associated OSA interface to terminate will dynamically terminate the HSCI. The user can optionally manually stop and restart the HSCI.

QDIO Accelerator

The QDIO Accelerator function extends the HiperSockets Accelerator function. HiperSockets Accelerator provides accelerated forwarding at the DLC layer for the following types of packets:

- Inbound packets over HiperSockets that are forwarded outbound over OSA-Express QDIO
- Inbound packets over OSA-Express QDIO that are forwarded outbound over HiperSockets

For more information about HiperSockets Accelerator, see "Efficient routing using HiperSockets Accelerator" on page 84.

QDIO Accelerator provides all of the function supported by HiperSockets Accelerator, and also provides accelerated forwarding at the DLC layer for the following types of packets:

- Inbound packets over OSA-Express QDIO that are forwarded outbound over OSA-Express QDIO
- Inbound packets over HiperSockets that are forwarded outbound over HiperSockets
- Sysplex distributor packets that are forwarded to a target stack, when the route involves any of the following inbound and outbound DLC combinations:
  - Inbound over HiperSockets, forwarded outbound over OSA-Express QDIO
  - Inbound over OSA-Express QDIO, forwarded outbound over HiperSockets
  - Inbound over OSA-Express QDIO, forwarded outbound over OSA-Express QDIO
  - Inbound over HiperSockets, forwarded outbound over HiperSockets

QDIO Acceleration is supported with or without the VIPAROUTE statement. When QDIO Accelerator is active, the stack dynamically creates QDIO Accelerator routes as it forwards packets in any of the inbound and outbound DLC combinations previously described. The DLC layer can perform accelerated

routing for packets across these routes, bypassing the IP forwarding function in the stack. Similarly, the stack dynamically creates QDIO Accelerator routes for packets that would be forwarded by the sysplex distributor in any of the inbound and outbound DLC combinations. The DLC layer can perform accelerated sysplex distributor routing for such packets.

To configure QDIO Accelerator, specify the QDIOACCELERATOR parameter on the IPCONFIG statement. For more information about the IPCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

You can display the QDIO Accelerator routing entries that are dynamically created for non-sysplex distributor packets using the Netstat ROUTE/-r report option with the QDIOACCEL modifier. For more information about the Netstat ROUTE/-r report, see z/OS Communications Server: IP System Administrator's Commands.

You can use the Netstat VCRT/-V report with the DETAIL modifier to display whether a sysplex distributor connection is eligible for acceleration. For more information about the Netstat VCRT/-V report, see z/OS Communications Server: IP System Administrator's Commands.

You can also use VTAM tuning statistics to monitor and measure the accelerated packets. For more information about gathering tuning statistics, see z/OS Communications Server: SNA Network Implementation Guide.

Restrictions:

- QDIO Accelerator is supported for IPv4 only.
- If IP security is enabled on the stack, the stack monitors IP filter rules and defensive filter rules that apply to routed traffic. Depending on your filter configuration, QDIO Accelerator might be restricted to only packets that are forwarded by the sysplex distributor. For more information about QDIO Accelerator and IP security, see "QDIO Accelerator and IP security" on page 102.
- If IP forwarding is disabled on the stack, then QDIO Accelerator applies only to packets that are forwarded by the sysplex distributor.
- Packets from the sysplex distributor to the target are not accelerated with the VIPAROUTE destination when the outbound interface is HiperSockets.

Tip: If packet trace is enabled, then packets that are accelerated by QDIO Accelerator do not appear in the packet trace (either inbound or outbound). If you want function similar to that of the packet trace in conjunction with QDIO Accelerator, you can use the OSA-Express network traffic analyzer (OSAENTA) to trace these packets that are accelerated to or from OSA-Express QDIO. For more details on OSAENTA, see "OSA-Express network traffic analyzer trace" on page 104.

QDIO Accelerator and IP security

When you enable IP security, the TCP/IP stack applies your configured filter policy to IP packets that it receives and sends. Depending on your filter policy, the IP layer might permit packets, deny packets, or protect packets using IPSec. For more information about IP security, see Chapter 17, "IP security," on page 883.

QDIO Accelerator processes some packets at the DLC layer, bypassing the IP layer. If such packets were allowed to be routed by QDIO Accelerator, this would prevent the IP layer from performing any special processing for these packets that is required by your filter policy.

When IP security is enabled, sysplex distributor packets are always eligible for acceleration using QDIO Accelerator because the packets are subject to IP filtering at the target stack rather than the distributor stack. However, QDIO Accelerator might not be able to forward routed traffic when IP security is enabled if any of your routed traffic is subject to special processing in your filter policy.

QDIO Accelerator forwards routed traffic only if your IP filter policy and defensive filters explicitly permit all routed traffic without logging. If your IP filter policy and defensive filters do not explicitly permit all routed traffic without logging, QDIO Accelerator does not forward routed traffic and one or more of the following messages is displayed on your console:

- EZD2020A

TCP/IP issues message EZD2020A if you are currently using the IP filters defined in your TCP/IP profile and one of the following conditions is true:

- There is no IPSECRULE statement with the ROUTING value `Routed` or `Either` to permit routed traffic.
- The first IPSECRULE statement that permits routed traffic does not permit all source addresses, all destination addresses, all protocols, and all security classes.
- The first IPSECRULE statement that permits all routed traffic specifies that logging is enabled, or logging is enabled as the default.

- EZD2021A

  TCP/IP issues message EZD2021A if you are currently using the IP filters defined in your policy configuration and one of the following conditions is true:

  - There is no IP filter rule that permits routed IPv4 traffic.
  - The first IP filter rule that applies to routed IPv4 traffic does not permit all IPv4 source addresses, all IPv4 destination addresses, both directions, all protocols, and all security classes.
  - The first IP filter rule that permits all routed IPv4 traffic specifies that logging is enabled, or logging is enabled as the default.

- EZD2022A

  TCP/IP issues message EZD2022A if you are using the Defense Manager daemon (DMD) to manage defensive filter rules, and you have a defensive filter installed that applies to routed traffic.

If you have enabled IP security and you want to allow QDIO Accelerator to forward routed traffic, see "Steps to allow QDIO Accelerator to forward routed traffic when IP security is enabled" on page 103.

Steps to allow QDIO Accelerator to forward routed traffic when IP security is enabled

Before you begin

Discuss all changes to your filter policy and defensive filters with your security administrator. Do not permit all routed traffic without first verifying that your network security policy specifies that all routed traffic is to be permitted.

Procedure

If you enabled IP security and you want to allow QDIO Accelerator to forward routed traffic, perform the following steps:

1. Consult your network security policy to determine how TCP/IP processes routed traffic.

   You cannot use QDIO Accelerator for routed traffic when any of the following conditions are true:

   - Some routed traffic must be denied in your IP filter policy.
   - Some routed traffic must be protected by IPsec in your IP filter policy.
   - Some routed traffic must be logged using TRMD.

   If any of these conditions are true, QDIO Accelerator forwards only sysplex distributor traffic.

2. If TCP/IP issues message EZD2020A and your network security policy specifies that all routed traffic is permitted, configure your TCP/IP profile to permit all routed traffic without logging:

   a) Ensure that the first IPv4 IPSECRULE statement with the Routing value `Routed` or `Either` permits all IPv4 addresses, all protocols, and all security classes.

      Tip: If your rule has the ROUTING value `Either`, the rule applies to both local and routed traffic. If your security policy does not allow you to permit all local traffic, split this rule into two rules, one with the Routing value `Routed` and one with the Routing value `Local`.

   b) Ensure that the first IPv4 IPSECRULE statement does not specify `LOG` to enable filter logging.

3. If TCP/IP issues message EZD2021A and your network security policy specifies that all routed traffic is permitted, configure your policy filter rules to permit all routed traffic without logging.
   a) If you are using the IBM Configuration Assistant for z/OS Communications Server to configure your IPSec policy, use the following settings:
      i) Ensure that the first connectivity rule that applies to routed IPv4 traffic specifies a topology of filtering only, applies to all IPv4 addresses, and uses a requirement map that maps all IP protocols and all security classes to a `Permit` security level.

         Tip: Your connectivity rule might apply to both local and routed traffic. If your security policy does not allow you to permit all local traffic, split this rule into two rules, one that applies to filtering for routed traffic and one that applies to filtering for local traffic.
      ii) Ensure that the first connectivity rule specifies that filter matches are not logged.
   b) If you are manually configuring your IPSec policy, use the following settings:
      i) Ensure that the associated IpService statement for the first IpFilterRule statement has the Routing value `Routed` or `Either`, permits all IPv4 addresses, all protocols, and all security classes, and has the Direction value `Bidirectional`.

         Tip: If your rule has the Routing value `Either`, the rule applies to both local and routed traffic. If your security policy does not allow you to permit all local traffic, split this filter rule into two filter rules, one with the Routing value `Routed` and one with the Routing value `Local`.
      ii) Ensure that the associated IpGenericFilterAction statement for the first IpFilterRule statement specifies the IpFilterLogging setting No to disable filter logging.
4. If TCP/IP issues message EZD2022A, and your network security policy specifies that all routed traffic is permitted, ensure that your defensive filters permit routed traffic:
   a) Issue the **`ipsec -F display`** command to display the defensive filters.
   b) If any filters are listed that apply to routed traffic, you have the following options:
      - Because defensive filters are temporary filters designed to address temporary conditions, you can wait for these filters to expire. After the filters expire, QDIO Accelerator can resume forwarding routed traffic.
      - If you and your security administrator determine that these defensive filters are no longer needed, your security administrator can issue the **`ipsec -F delete`** command to delete these filters so that QDIO Accelerator can resume forwarding routed traffic.

   For more information about the **`ipsec`** command, see z/OS Communications Server: IP System Administrator's Commands.

OSA-Express network traffic analyzer trace

When data problems occur in a network with OSA adapters, multiple traces are usually required. A sniffer trace might be required to see the data as it was received from or sent to the network, an OSA hardware trace might be required if the problem is suspected in the OSA, and z/OS Communications Server traces are required to diagnose VTAM or TCP/IP problems.

To help with problem diagnosis, the OSA-Express network traffic analyzer (OSAENTA) function provides a way to trace inbound and outbound frames for an OSA-Express2 or later feature in QDIO mode. The OSAENTA trace function is controlled and formatted by z/OS Communications Server, but is collected in the OSA at the network port. You can control the OSAENTA trace function by using either the OSAENTA statement in the TCP/IP profile or the VARY TCPIP,,OSAENTA command. The controls that are provided include the ability to filter what data is collected by parameters such as IP address, TCP/UDP port, or frame type, and to specify how much data is to be collected. They also provide the capability to trace frames that are discarded by the OSA-Express2 or later feature. You can display current settings for the OSAENTA trace function by using the Netstat DEvlinks/-d command.

Because the data is collected at the Ethernet frame level, you can use this function to trace the MAC headers for packets, a capability that is not provided by existing packet traces. It also enables the tracing of other types of packets that existing packet traces do not contain, including the following packets:

- ARP packets
- Packets to and from other users sharing the OSA, including other TCP/IP stacks, z/Linux users, and z/VM users
- SNA packets

There are obvious security considerations for this type of trace. Security control is enabled at the Hardware Management Console (HMC) of the OSA-Express2 or later feature. To trace packets for stacks or images other than the operating system image where you activate the OSAENTA trace interface, you must use the HMC to configure the OSA to enable this tracing.

The OSAENTA trace function communicates with the OSA-Express feature being traced by using a dynamically defined QDIO interface to the OSA-Express feature. The interface is created when the first VARY TCPIP,,OSAENTA command for that OSA is entered, or the first OSAENTA TCP/IP profile statement is processed. The interface is named EZANTA*xxxxxxxx*, where *xxxxxxxx* is the port name of the OSA specified on the VARY TCPIP,,OSAENTA command, and must match the PORTNAME parameter on the VTAM TRLE representing the traced OSA-Express feature. This also means that a TRLE must be defined in VTAM for this OSA-Express trace interface, though the same TRLE used for MPCIPA devices and interfaces is used for this dynamically defined trace interface. The EZANTA*xxxxxxxx* interface is used exclusively for receiving trace records from that OSA-Express feature. It is started, stopped, and deleted with the ON, OFF, and DEL parameters of the VARY TCPIP,,OSAENTA command.

When the OSAENTA trace is enabled, the received trace records are collected by Component Trace (CTRACE) using the SYSTCPOT component. The traced records can then be formatted using the IPCS CTRACE command, specifying the component name SYSTCPOT. For information about retrieving the OSA-Express network traffic analyzer data in real time, see z/OS Communications Server: IP Programmer's Guide and Reference.

The OSAENTA trace can have a negative impact on performance if sufficient trace filters are not specified before you enable the trace. OSAENTA can reduce the amount of traffic that the OSA-Express feature can process and the amount of traffic that can be accelerated through that OSA-Express feature. Also, host processing to collect the OSAENTA trace records can increase host CPU consumption. Specify sufficient filters to limit the amount of traffic that is traced to only what is necessary for problem diagnosis.

For more information about the OSAENTA statement, see z/OS Communications Server: IP Configuration Reference. For more information about the VARY TCPIP,,OSAENTA command, see z/OS Communications Server: IP System Administrator's Commands. For more information about formatting an OSAENTA trace, see z/OS Communications Server: IP Diagnosis Guide. For the level of OSA-Express2 or later feature that supports OSAENTA, see the 2094DEVICE Preventive Service Planning (PSP) bucket and the 2096DEVICE Preventive Service Planning (PSP) bucket.

Synchronization of OSA-Express2 or later diagnostic data

VTAM provides the externals that control the QDIOSYNC trace facility, used to synchronize OSA-Express2 or later diagnostic data with host diagnostic data. For information about the use of the QDIOSYNC trace facility, see z/OS Communications Server: SNA Diagnosis Vol 1, Techniques and Procedures.

Prioritizing outbound OSA-Express data using the Workload Manager service class

The z/OS Workload Manager (WLM) provides a priority associated with each unit of work that runs a TCP/IP socket API to send data. The priority provided by WLM is related to the WLM service class associated with the unit of work. The unit of work can derive its priority from the service class associated with the address space in which it is running or from the service class associated with the enclave to which it belongs. The priorities provided by WLM, from highest priority to lowest priority, are as follows:

- System-defined service class (SYSTEM), used for system address spaces
- System-defined service class (SYSSTC), used for high-priority started tasks
- User-defined service classes with importance level 1
- User-defined service classes with importance level 2

- User-defined service classes with importance level 3
- User-defined service classes with importance level 4
- User-defined service classes with importance level 5
- User-defined service classes associated with a discretionary goal

For more information about WLM and the WLM service classes, see z/OS MVS Planning: Workload Management and z/OS MVS Programming: Workload Management Services.

z/OS Communications Server supports four priority values for outbound QDIO traffic, 1 through 4, with 1 being the highest priority. Priorities with lower numbers are given preferential treatment by the QDIO device driver and the OSA-Express feature.

You can specify the QDIO priority value to be used on the SetSubnetPrioTosMask statement in a Quality of Service (QoS) policy, which provides one way to influence the QDIO traffic priority that is used for outbound packets. The SetSubnetPrioTosMask statement maps the IPv4 type of service (ToS) byte or IPv6 traffic class to these four QDIO traffic priorities. For more information about the SetSubnetPrioTosMask statement, see z/OS Communications Server: IP Configuration Reference.

Another way to influence the QDIO traffic priority is to use the WLMPRIORITYQ parameter on the GLOBALCONFIG profile statement. The WLMPRIORITYQ parameter automatically extends the preferential treatment of the most important workloads for a business through the QDIO device driver all the way to the LAN. When the WLMPRIORITYQ parameter is specified and a packet with a ToS or traffic class value 0 is sent over an OSA-Express feature in QDIO mode, Communications Server sets the OSA-Express write priority of the packet based on the priority value provided by the WLM service class. In addition, the WLMPRIORITYQ parameter can be used to influence the QDIO traffic priority of forwarded packets with a ToS or traffic class value 0.

For more information about the GLOBALCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

Restrictions:

- Prioritization using the WLM service class is effective only when enabled and when the ToS or traffic class value is 0.
- Prioritization using the WLM service class is ineffective for interfaces other than OSA-Express features in QDIO mode.
- Prioritization of forwarded packets is ineffective unless DATAGRAMFWD is specified on the IPCONFIG statement, the IPCONFIG6 statement, or both statements.
- The WLMPRIORITYQ setting for forwarded packets has no effect on accelerated packets. To set the write priority for accelerated packets, use QDIOPRIORITY on the IQDIOROUTING parameter or the QDIOACCELERATOR parameter on the IPCONFIG profile statement. For more information about the IPCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

Fixed storage requirements for OSA-Express QDIO and HiperSockets interfaces

Each OSA-Express QDIO and HiperSockets interface requires fixed storage for read processing (which is allocated by VTAM). If you define a large number of these interfaces (for example, by configuring multiple VLANs to one or more OSA-Express features), then you need to consider how much fixed storage your configuration requires.

For information about how much fixed storage VTAM allocates by default for each OSA-Express QDIO and HiperSockets interface, how to control the amount of this storage allocation using the VTAM QDIOSTG start option (for OSA-Express QDIO) and the VTAM IQDIOSTG start option (for HiperSockets), and considerations for the IVTPRM00 parmlib member, see z/OS Communications Server: SNA Resource Definition Reference.

You can also override the global QDIOSTG or IQDIOSTG value and control the amount of fixed storage for a specific OSA-Express QDIO or HiperSockets interface by using the READSTORAGE parameter on the LINK and INTERFACE statements.

Using TEMPIP interfaces

You can configure a TEMPIP interface by configuring the interface with the TEMPIP parameter. The TEMPIP parameter allows the interface to be started with an IP address of 0.0.0.0 instead of a statically defined IP address. The 0.0.0.0 IP address enables TEMPIP interfaces to broadcast and multicast IP traffic. Unicast traffic cannot use TEMPIP interfaces. The TEMPIP parameter can be configured only on the IPv4 IPAQENET OSA-Express QDIO INTERFACE statement.

About this task

TEMPIP interfaces are used in a unit test environment to support applications that provide a DHCP client, such as IBM Rational® Developer for System z Unit Test feature (Rdz-UT).

Procedure

To use a TEMPIP interface, take the following steps:

1. Define an interface with the TEMPIP parameter instead of the IPADDR parameter.
2. Start the interface.
3. Use UDP broadcasts to communicate with the DHCP server to obtain an IP address, subnet, default router, and DNS servers.
4. Stop the interface.
5. Use the V TCPIP,,OBEYFILE command to delete the interface.
6. Configure the interface in the TCP/IP profile with the IPADDR parameter and the IP address that is obtained with DHCP. Routing information and DNS server information must be configured in the correct data files. Use the V TCPIP,,OBEYFILE command to add the new interface.
7. Start the interface.

Guidelines for using TEMPIP interfaces

Follow these guidelines when you are using TEMPIP interfaces:

- Do not code the SOURCEVIPAINTERFACE parameter because DHCP servers expect a source IP address of 0.0.0.0 on packets from the DHCP client.
- If the VMAC parameter is defined on the INTERFACE statement, you must specify the virtual MAC address. If you delete a device and then add it back, the generated VMAC address of this device is changed.
- You must define the IPBCAST parameter on the INTERFACE statement to allow broadcasts.
- The VLANID parameter is supported for TEMPIP interfaces, except when multiple interfaces are sharing an OSA-Express feature.

Determining the maximum transmission unit

TCP/IP uses the maximum transmission unit (MTU) value to determine the largest sized frame to send. The MTU value that is in effect for a given outbound send is one of the following two values:

- Path MTU value

  TCP/IP automatically enables path MTU discovery for IPv6. If a packet is an IPv6 packet, or if a packet is an IPv4 packet and path MTU discovery is enabled, the path MTU value is used to determine the maximum size of the packet. Path MTU discovery initially sets the path MTU value to the actual route MTU value for the route. If packets require fragmentation to get to the final destination, path MTU discovery determines the path MTU value by repeatedly decreasing the value until it can send packets to the final destination without fragmentation.

  Guideline: You can enable path MTU discovery for IPv4 by configuring IPCONFIG PATHMTUDISCOVERY in the TCP/IP profile.
- Actual route MTU value

The actual route MTU value is the lesser of the interface MTU value and the configured route MTU value. If path MTU discovery is not enabled, the actual route MTU value is used.

– Interface MTU value

  The interface MTU value is a characteristic of an interface, and is either learned from the device during activation or is hardcoded based on the type of the physical device. For information about the interface MTU values that TCP/IP uses for the various network interface types supported by TCP/IP, see the summary of DEVICE and LINK statements and the summary of INTERFACE statements in z/OS Communications Server: IP Configuration Reference. For an IPAQENET6 interface, or for IPAQENET or IPAQIDIO interfaces defined with the INTERFACE statement, you can configure a lower interface MTU value using the MTU keyword on the INTERFACE statement.

  Restriction: You cannot modify the interface MTU value for IPAQENET or IPAQIDIO interfaces defined using the DEVICE, LINK, and HOME statements.

  Results:

  - The TCP/IP stack sets the interface MTU value to the lesser of the learned MTU value and the MTU value configured on the INTERFACE statement.
  - For an active link or interface, TCP/IP reports the interface MTU value in the ActMtu field of the Netstat DEVLINKS/-d report.

– Configured route MTU value

  The configured route MTU value is the MTU size that is configured for a route.

  - Static route

    For a static route, you can specify the configured route MTU value in the TCP/IP profile on a ROUTE entry in a BEGINROUTES block.

  - IPv4 dynamic route

    For IPv4 dynamic routes over an interface that are added by OMPROUTE, the configured route MTU value is the value of the MTU keyword specified on the RIP_INTERFACE, OSPF_INTERFACE or INTERFACE statement in the OMPROUTE configuration file for the outgoing interface of the route.

    Result: If you do not specify an MTU value for an interface in the OMPROUTE configuration file, OMPROUTE uses the value 576.

  - IPv6 dynamic route

    For IPv6 dynamic routes added by OMPROUTE, OMPROUTE learns the interface MTU value from TCP/IP; you cannot configure a route MTU value in the OMPROUTE configuration file.

    Result: For IPv6 dynamic routes that are learned by OMPROUTE, the configured route MTU size is the same as the interface MTU size.

These factors comprise a general set of rules for how TCP/IP determines the MTU, but there are some exceptions. For example, if an application uses the IPV6_USE_MIN_MTU socket option, TCP/IP sends outbound packets using the IPv6 minimum MTU value 1280.

Guidelines:

- Enable path MTU discovery in configurations where traffic originating in the z/OS TCP/IP stack will traverse multiple hops with different MTU sizes.
- The OSA-Express Gigabit Ethernet adapter supports an interface MTU value of 8992, but not all routers and switches support an MTU value of this size. If you are using this adapter and any routers or switches in your configuration do not support an MTU value of 8992, then either configure a lower MTU value for your routes or specify a lower MTU value on the INTERFACE statement in the TCP/IP profile.
- When you are using OMPROUTE, specify the MTU keyword for each IPv4 interface.
- When you are using OMPROUTE, configure all nodes on a LAN to use the same MTU value. Otherwise, you might encounter problems, such as OSPF adjacency errors.

Considerations for multiple servers sharing a TCP port

For a TCP server application to support a large number of client connections on a single system while providing good performance for those connections, it might be necessary to run more than one instance of the server application to service the connection requests. For all instances of the server application to receive client connection requests without changing the client applications, the servers must all bind to the same server IP address and port. To enable this in the TCP/IP stack, you must add the SHAREPORT or SHAREPORTWLM keyword to the PORT profile statement that reserves the TCP port for the server instances. For more detailed information on the PORT statement and its keywords, see z/OS Communications Server: IP Configuration Reference.

The set of server instances sharing the same TCP port on the same TCP/IP stack is called a shareport group. As incoming client connections arrive for this port and IP address, TCP/IP distributes them across the servers in the shareport group.

If the SHAREPORT keyword was specified, the connections are distributed across the available servers using a weighted round-robin distribution based on the Servers' accept Efficiency Fractions (SEFs).

If the SHAREPORTWLM keyword was specified, the connections are distributed across the available servers using a weighted round-robin distribution based on the WLM server-specific recommendations, modified by the SEFs.

The SEF is a measure, calculated at intervals of approximately 1 minute, of the efficiency of the server application in accepting new connection requests and managing its backlog queue. You can use the Netstat ALL/-A command to obtain the current SEF value for a server. For more detailed information about the Netstat command, see z/OS Communications Server: IP System Administrator's Commands.

Configuring a shareport group can also be used to improve the availability of key applications. For example, should an application server instance become inactive (encounters a failure or is stopped for planned maintenance), the other applications in the shareport group can continue to process client requests.

Applications that might be good candidates for being placed in a shareport group must also satisfy the following requirements:

- Multiple instances of the application can be started within a single system.
- Each application instance must provide the same functional capabilities. That is, each must be capable of processing any TCP connection requests to the shared port.

In addition, each TCP connection directed to a shareport group must be eligible for load balancing (it cannot have an affinity to a specific server). If specific client affinities exist to a specific server, you should consider using sysplex distributor, which provides for load balancing of connections while maintaining client affinities to a specific server instance over a period of time (Timed Affinity). While the primary focus of the sysplex distributor is on load balancing across servers on multiple target systems, it also supports multiple servers in a shareport group on the target systems. For more information on sysplex distributor configuration, see Chapter 8, “TCP/IP in a sysplex,” on page 435.

Considerations for Common Information Model providers

The Common Information Model (CIM) provides a model for describing and accessing data across an enterprise. CIM is a standard developed by a consortium of independent hardware and software vendors, including IBM, called the Distributed Management Task Force (DMTF). CIM data is defined by standardized CIM classes. Class definitions define properties for each class. For example, some properties of the CIM_EthernetPort (Ethernet interface) class are interface type, enabled state, and speed. Platforms can extend the standardized CIM classes by defining their own platform-specific classes and their properties.

CIM data instrumentation is supplied by CIM components called providers. The providers gather data on a system in support of the CIM classes. Clients can retrieve the data through the Common Information Model Object Manager. On z/OS, this function is provided by the z/OS CIM server.

Communications Server supplies CIM providers for the CIM classes shown in Table 8 on page 110. Only IPv4 data is currently supported. To support data from eight TCP/IP stacks on an MVS image, the TCP/IP stack name has been added to the IBMzOS_EthernetPort and IBMzOS_IPProtocolEndpoint classes as a platform-specific property.

For more information about CIM, the z/OS CIM server, the properties support for the Communications Server CIM classes, and other z/OS CIM provider support, see z/OS Common Information Model User's Guide.

Table 8. Communications Server CIM providers

CIM base class name	z/OS class name	Description
CIM_EthernetPort	IBMzOS_EthernetPort	IPv4 Ethernet interfaces defined to the TCP/IP stack
CIM_IPProtocolEndpoint	IBMzOS_IPProtocolEndpoint	IPv4 IP addresses defined to the TCP/IP stack
CIM_PortImplementsEndpoint	IBMzOS_NetworkPortImplementsIPEndpoint	IPv4 Ethernet interfaces and their associated IP addresses
CIM_SystemDevice	IBMzOS_CSNetworkPort	MVS system and its associated IPv4 Ethernet interfaces

This CIM provider function is in the /usr/lpp/tcpip/lib directory. There is no configuration necessary to activate this CIM provider support. The z/OS CIM server must be configured and activated for the data supported by the Communications Server CIM providers to be available to clients. The z/OS Communications Server CIM classes are shipped with the z/OS CIM server. The files that define these classes and any platform specific properties are also installed in the /usr/lpp/tcpip/mof directory.

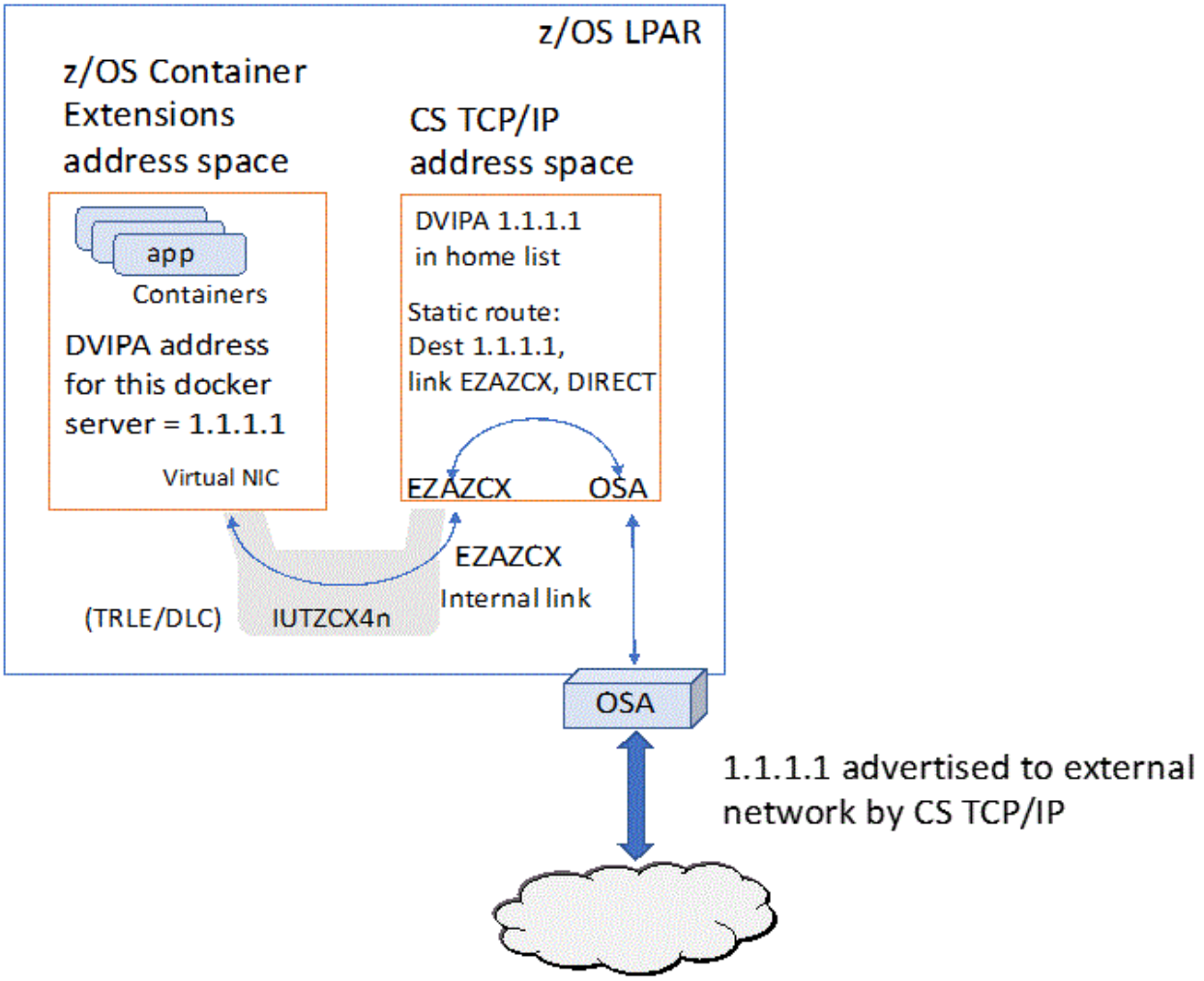
There are security configuration considerations that govern the ability of providers to gather CIM data. For more information about security configuration, see "CIM provider access control" on page 148.

IBM z/OS Container Extensions network overview

IBM z/OS Container Extensions (zCX) provides an execution environment allowing z/OS to host applications based on Linux that are managed with docker containers. Each instance of zCX (unique zCX job) is provisioned within a unique z/OS address space. The zCX address space represents a virtual server which hosts applications managed in docker containers. The virtual server does not require operational tasks from the z/OS environment or the z/OS administrator. From an operational perspective, the virtual server is transparent to the z/OS environment. The zCX environment is configured using z/OSMF. For information about configuring the zCX environment, see IBM z/OS Container Extensions Guide and *IBM z/OS Management Facility Online Help* for Configuration Workflow.

z/OS Communications Server provides network communications and network related services for the zCX workloads. The Linux virtual server is represented by a unique type of application instance DVIPA called a zCX DVIPA. The VIPARange statement is used with the ZCX keyword to create zCX DVIPAs. Defining zCX DVIPAs is the primary IP configuration task and in many instances it will be the only required configuration task. For additional information about zCX DVIPAs, "Configuring application-instance DVIPAs for IBM z/OS Container Extensions (zCX)" on page 380.

The following figure provides an overview of the z/OS Communications Server support provided for the zCX environment.



Users define zCX DVIPAs to TCP/IP using VIPARange with the ZCX parameter using Network Configuration Assistant or directly in their TCP/IP profile.

The remaining steps are automatic:

1. EZAZCX interface is automatically created[1] by TCP/IP and connected to a dynamically created internal network represented by a dynamically created TRLE called IUTZCX4n. A unique instance of the IUTZCX4n DLC is created for each TCP/IP stack connecting to zCX servers. The n represents the instance of each IUTZCX4n DLC created.
2. When the zCX job is started by the user:
   a. this instance of zCX binds to the DVIPA and connects to TCP/IP over EZAZCX
   b. DVIPA (zCX) is activated / added to the home list
   c. TCP/IP creates a static route to the zCX DVIPAs for its own internal use. This route is local to and controlled by the stack that owns the zCX and is not advertised using dynamic routing.
3. When using dynamic routing the zCX DVIPAs are treated by OMPROUTE as application-instance DVIPAs and are advertised accordingly.[2]

[1] The EZAZCX interface is automatically created when the IUTSAMEH interface is created (either by DYNAMICXCF or static DEVICE/LINK/HOME for IUTSAMEH) and at least one VIPARANGE ZCX is configured. The EZAZCX interface transitions to ready when the first zCX DVIPA is activated. The EZAZCX interface is created for zCX instances using an IPv4 DVIPA. If the zCX instance is using an IPv6 DVIPA, the interface created is EZ6ZCX and the TRLE created is IUTZCX6n.

[2] When using dynamic routing, you must define the zCX DVIPAs to OMPRoute just like any other application-instance DVIPA. OMPROUTE will then advertise the zCX DVIPA when it activates on the host. When the

4. CS TCP/IP forwards packets for 1.1.1.1 over the zCX IP route and the EZAZCX Interface.

Note: IP filters can be configured and applied during IP forwarding. IPSec tunnels can be applied to the external IP routes. If you have IP filters defined, updates to your IP filter rules are required. You must ensure that you permit ROUTED and LOCAL (EITHER) traffic for the zCX DVIPAs.

z/OS Container Extensions IPv6 network overview

z/OS supports zCX instances with both IPv4 and IPv6 connectivity. A zCX instance can support just IPv4 or both IPv4 and IPv6 connectivity.

Defining zCX DVIPAs using VIPARange remains as the primary IP configuration task that controls and enables each zCX instance for each specific IP version. When the zCX job is started, the appropriate zCX interfaces will be started.

Configuring and enabling zCX IPv6 requires that your zCX instance is already enabled for IPv4 (see the above IPv4 steps) and then the required IPv6 steps are summarized below for the following two use cases:

- **z/OS TCP/IP users who already have enabled z/OS for IPv6 (key steps):**
  1. Define IPv6 zCX DVIPAs using the VIPARange statement.

     Note:

     a. A separate IPv6 VIPARange statement should be created for each zCX instance that requires IPv6 connectivity.

     b. For supporting high availability for zCX, your IPv6 VIPARange statements should be configured in all z/OS TCP/IP instances within the sysplex eligible to host this zCX IPv6 instance.
  2. Specify the IPv6 DVIPA in the z/OSMF zCX workflows (provisioning or reconfiguration workflows). Optionally specify any other IPv6 addresses or hostnames for any zCX configuration options, such as DNS, Registry, Proxy, or LDAP addresses using the z/OSMF zCX workflows.
  3. Common IPv6 configuration steps:

     Most existing z/OS IPv6 users will have already completed the following common IPv6 steps, but here are some key steps to consider:

     a. External IPv6 Interfaces:

        To enable IPv6 communications with hosts external to this z/OS TCP/IP instance you must enable the associated z/OS IPv6 interfaces, such as OSA or HiperSockets.

     b. IPv6 Dynamic XCF must be enabled.[3]

     c. If using dynamic routing, define your IPv6 DVIPAs to OMPRoute.

        Note: IP filters can be configured and applied during IP forwarding. IPSec tunnels can be applied to the external IPv6 routes. If you have IP filters defined, updates to your IP filter rules are required. You must ensure that you permit ROUTED and LOCAL (EITHER) traffic for the zCX DVIPAs.
- **z/OS TCP/IP users who have not enabled z/OS for IPv6:**
  1. This type of user must first complete the z/OS IPv6 migration or enablement. This type of user must start with the z/OS Communications Server: IPv6 Network and Application Design Guide.

DVIPA is moved, external hosts will automatically (with dynamic routing updates) find the new location of the DVIPA and the related zCX applications. If OMPROUTE is not being used, you must ensure the zCX DVIPAs can be reached by other hosts by defining static routes on other hosts that need to reach the zCX instance.

[3] The EZAZCX interface is automatically created when the IUTSAMEH interface is created (either by IPCONFIG(6) DYNAMICXCF or static DEVICE/LINK/HOME for IUTSAMEH) and at least one VIPARANGE ZCX is configured. The EZAZCX interface transitions to ready when the first zCX DVIPA is activated. The EZAZCX interface is created for zCX instances using an IPv4 DVIPA. If the zCX instance is using an IPv6 DVIPA, the interface created is EZ6ZCX and the TRLE created is IUTZCX6n.

2. Once your z/OS TCP/IP stack is enabled for IPv6, review the first list above for the specific zCX IPv6 steps.

Required steps before starting TCP/IP

This information describes the steps you must complete before starting TCP/IP.

Planning your installation and migration

It will be to your advantage to have thoroughly studied the following documentation prior to the installation and customization of z/OS Communications Server:

- Program Directory for z/OS for CBPDO Installation and ServerPac Reference, Program Number 5650-ZOS
- Preventive Service Planning (PSP) bucket
- z/OS Communications Server: New Function Summary
- z/OS UNIX System Services Planning
- OS390CKL, IBM MKTTOOLS information for the z/OS UNIX System Services implementer

It is also recommended that you attend a z/OS UNIX System Services concepts class and a class in using z/OS UNIX System Services prior to migrating to z/OS Communications Server. If this is not possible, then you will want to ensure that the z/OS UNIX System Services implementer and the RACF administrator work together with you during the installation and customization process.

Planning for and installing z/OS Communications Server requires MVS, UNIX, and networking skill. If your background is in traditional MVS programming or system programming, the z/OS UNIX System Services terminology might at first seem to be somewhat confusing. If your background is in the UNIX environment, the terms should be familiar to you.

In the past, MVS TCP/IP system programmers have needed a working knowledge of the MVS or z/OS system. These programmers have been accustomed to working closely with the RACF administrator and z/OS system programmer for authorizations; the IP address administrator for basic name and address assignments; and the administrators of the router network and channel-attached peripherals for connection definition and problem determination.

With the introduction of z/OS Communications Server, the TCP/IP system programmer needs to develop an additional alliance with the z/OS UNIX System Services system programmer. The TSO interfaces that have been traditionally available in the host-based TCP/IP still stand at the system programmer's disposal and additional MVS console commands simplify some TCP/IP operations. However, another user interface provided by the UNIX shell environment, either with the z/OS shell or the ISPF SHELL, is a useful and sometimes necessary tool that the TCP/IP system programmer will need to work with. Additionally, the tight coupling of z/OS Communications Server with z/OS UNIX System Services means that the TCP/IP system programmer needs more than a passing knowledge of UNIX conventions, commands, and hierarchical file system concepts. Even if the system programmer is familiar with other UNIX environments, work with the UNIX shell requires more than basic familiarity.

In the first version of a full TCP/IP stack based on native MVS and on z/OS UNIX System Services, few have all the requisite skills to successfully implement z/OS Communications Server on their own. As more and more system programmers acquire skills in UNIX System Services and in TCP/IP, this will become less and less the case. Working with the z/OS UNIX System Services implementer when implementing z/OS Communications Server provides the most effective solution to establishing a working z/OS Communications Server environment.

If you are migrating to z/OS Communications Server, establish a migration process to move all your existing applications, and after this, consider the use of new and enhanced functions based on z/OS Communications Server: New Function Summary. z/OS Communications Server allows multiple copies of the TCP/IP protocol stack to execute on the same MVS image. However, with all the performance enhancements introduced in z/OS Communications Server, it is probably not necessary to implement a multi-stack system for production purposes unless one is considering building a system programming test stack.

Step 1: Install z/OS Communications Server

Before you begin the installation:

- To help you plan the installation and migration of z/OS Communications Server, see:
  - z/OS Planning for Installation
  - z/OS Communications Server: New Function Summary
  - z/OS Upgrade Workflow
- Be sure you understand the data set naming conventions used in TCP/IP. You can find this information in "Configuration data set naming conventions" on page 13.
- Consult the *z/OS Program Directory* (Customization considerations for Wave 1D) for current information about the material, procedures, and storage estimates of the MVS image.

Install z/OS Communications Server with other elements of z/OS. If you use the ServerPac method of installation, see *z/OS Installing Your Order*; if you use the CBPDO method of installation, see *z/OS Program Directory*. When appropriate, that information will direct you back to this information to customize the TCP/IP data sets and procedures and verify their configuration.

Verifying the initial installation

Both the *z/OS Program Directory* and *z/OS Installing Your Order* contain step-by-step instructions that can be used to set up and verify a basic TCP/IP configuration with only the loopback address and a few key servers. For more information regarding these instructions, see the information about Wave 1D customizations in the z/OS Program Directory or the information about verifying your installation in *z/OS Installing Your Order*.

Step 2: Customize z/OS Communications Server

To customize TCP/IP you need to update the cataloged procedures and configuration data sets for the TCP/IP address space, its clients, and servers.

z/OS Communications Server runs as a started task in its own address space. Each of the servers runs in its own address space and is started with its own procedure. The TCP/IP address space requires:

- A procedure in a system or recognized PROCLIB.
- A data set that provides configuration definitions for the TCP/IP address space and includes statements affecting many of the servers. This data set is referred to as PROFILE.TCPIP.
- A data set to provide the parameters that are common across all clients. This data set is referred to as TCPIP.DATA.

Other topics show you how to:

- Configure the TCP/IP address space by updating the samples provided in SEZAINST(SAMPPROF) and SEZAINST(TCPIPROC).
- Configure the universal client parameters provided in SEZAINST(TCPDATA).
- Configure the site table, defined in *hlq*.HOSTS.LOCAL or *hlq*.ETC.IPNODES, to identify the Internet names and addresses of your TCP/IP host.
- Customize the TCP/IP Component Trace parameters by updating the CTRACE parameter in the PARM= field of the EXEC JCL statement in the TCP/IP started procedure.

  You can find a description of the MVS Component Trace support in the z/OS Communications Server: IP Diagnosis Guide.
- Specify the ENVAR parameter on the PARM=keyword to override the resolver file. For more information on setting the environment variable RESOLVER_CONFIG using the ENVAR parameter, see "Considerations for multiple instances of TCP/IP" on page 45.
- Configure each of the servers you want to run. This might require:
  - Modifying sample procedures and adding them in your PROCLIB

- Modifying the configuration data set, PROFILE.TCPIP
- Adding port numbers to *hlq*.ETC.SERVICES
- Modifying other data sets containing server-specific parameters

You can find the sample procedures and data sets in SEZAINST. Table 1 on page 15 provides additional reference information you can use as you configure and customize each server.

You can find general information about starting, stopping, and dynamically controlling the servers in z/OS Communications Server: IP System Administrator's Commands.

Many of the servers also require other data sets for their specific functions.

Making SYS1.PARMLIB changes

You need to make certain changes to SYS1.PARMLIB. These changes depend on which of the following installation methods you use:

ServerPac method
After the file system is restored (through the RESTFS job), you will see that ServerPac has changed some of the parmlib members. Follow the instructions to change the BPXPRMxx member of parmlib.

CBPDO method
Change the parmlib members according to the instructions listed in the installation instructions for Wave 1. Tables describing changes to parmlib and changes to the BPXPRMxx member are included.

Note: z/OS Communications Server exploits z/OS UNIX services even for traditional MVS environments and applications. Before using TCP/IP services, the z/OS UNIX environment must be set up in full function mode. For a list of tasks involved in setting up for full function mode, see z/OS UNIX System Services Planning. System Managed Storage (SMS, which is part of the DFSMSdfp element of z/OS) must be configured, and you will need to customize SMS, RACF, and the z/OS UNIX file system.

Additional information about required TCP/IP definitions for the UNIX environment can be found in "Defining TCP/IP as a UNIX System Services physical file system" on page 39 and "UNIX System Services security considerations" on page 36.

Common z/OS UNIX configuration problems

The following list includes some explanations and possible solutions for common problems that you might encounter when you are configuring the z/OS UNIX environment.

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIP-BPX1SOC,
     00000003,FFFFFFFF,00000070,112B00B6
```

  These messages usually indicate that both INET and CINET FILESYSTYPE have been specified. Only one should be specified; see the FILESYSTYPE information in z/OS UNIX System Services Planning for additional information.

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIP-BPX1SOC,
     00000003,FFFFFFFF,0000006F,112B00B0
```

  These messages indicate that the requester of the service is not privileged. The service requested requires a privileged user. Check the documentation for the service to understand what privilege is required.

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR
     FOR TCPIPA-BPX1IOC,8008C981,FFFFFFFF,0000009E,12B2005A

EZZ4204I TCPIP INITIALIZATION FOR TCPIPA FAILED
```

These messages usually indicate that an incorrect jobname was specified in the SUBFILESYSTYPE NAME() definition in the BPXPRM*xx* member for a common INET environment. In this scenario, the NAME() must match TCPIPA.

- TCP/IP initialization fails with the following messages:

```
IEA8481 DUMP SUPPRESSED - ABDUMP MAY NOT DUMP STORAG FOR KEY 0-7 JOB TCPV34A
IEF4501 TCPIPA TCPIPA - ABEND=SEC6 U0000 REASON=0F01C008
```

These messages are usually an indicator that an OMVS RACF segment has not been defined for the user ID associated with the TCP/IP started procedure. Define an OMVS segment with a UID of 0 for the user ID associated with the TCP/IP started procedure.

- TCP/IP initialization fails with the following messages:

```
IEF4031 TCPIPA - STARTED - TIME=16.01.25
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIPA-BPX11OC,
     8008139A,FFFFFFFF,00000079,12D2025E
EZZ4204I TCPIP INITIALIZATION FOR TCPIPA FAILED.

==> The 0079 value is EINVAL - The parameter is incorrect
==> The 025E value is JRSocketCallParmError - A socket syscall
    contains incorrect parameters
```

These messages usually indicate that an incorrect entry point name has been specified in the SUBFILESYSTYPE ENTRYPOINT() definition. The correct value is ENTRYPOINT(EZBPFINI).

- TCP/IP initialization fails with the following messages:

```
EZZ4203I Z/OS UNIX-TCP/IP CONNECTION ERROR FOR TCPIPA-BPX1SOC,
     00000003,FFFFFFFF,0000045A,112B0000
EZZ4204I TCPIP INITIALIZATION FOR TCPIPA FAILED.

==> The 045A value is EAFNOSUPPORT - The address family is not supported
```

These messages indicate that AF_INET was not defined or did not initialize properly. Check for any earlier z/OS UNIX messages and verify that the z/OS UNIX NETWORK DOMAINNAME(AF_INET) statement is in your BPXPRM*xx* member.

- After issuing a NETSTAT command from TSO, the following message is displayed:

```
netstat
CEE5101C During initialization, the z/OS UNIX callable service
     BPX1MSS failed. The system return code was 0000000156,
     the reason code was 0507014D. The application will be
     terminated.
NETSTAT ENDED DUE TO ERROR+
READY
?
USER ABEND CODE 4093 REASON CODE 00000090
READY

==> The 0156 value is EMVSINITIAL - Process initialization error
==> The 014D value is JRFsFailChdir - The dub failed, due to
    an error with the initial home directory
```

These messages indicate that the user ID issuing the NETSTAT command does not have an OMVS RACF segment defined for it. Define an OMVS segment for this user ID or activate the automatic assignment of unique UNIX identities support. For details, see “UNIX System Services security considerations” on page 36.

- Socket applications using the z/OS Communications Server TCP/IP Services APIs fail with an ERRNO of 156.

  ERRNO 156 indicates a z/OS UNIX process initialization failure. This is usually an indication that an appropriate OMVS RACF segment is not defined for the user ID associated with the application. The

RACF OMVS segment may not be defined or may contain errors such as an improper HOME() directory specification. If the OMVS segment is not defined, you may also receive the following message:

```
ICH4081 USER(USER8   ) GROUP(SYS1   )   NAME(TSO USERID USER8  )
   CL(PROCESS )
   OMVS SEGMENT NOT DEFINED
```

In this example, USER8 is the user ID associated with the failing application. To correct this problem, define an appropriate OMVS segment for the user ID associated with the failing application. For details, see "UNIX System Services security considerations" on page 36.

Step 3: Configure VMCF and TNF

The Pascal socket interface uses the IUCV/VMCF services for a limited set of inter-address space communication flows. As a result, if you are using any applications (provided by IBM or others) that use the Pascal socket API, you must ensure that the Virtual Machine Communication Facility (VMCF) and Termination Notification Facility (TNF) subsystems are active before the applications are started. TCP/IP provides the following applications and commands that use the Pascal socket interface:

- LPD servers
- TSO HOMETEST, LPQ, LPR, LPRM, LPRSET, TELNET, and TESTSITE commands

If you are using any of these applications or commands, you need to set up VMCF and TNF.

You can configure VMCF and TNF in two different ways: as restartable subsystems or as non-restartable subsystems.

If you configure VMCF and TNF as restartable subsystems and you want the VMCF node name to be used as a default host name during TCP/IP initialization (when no other host name can be located), VMCF must be started before TCP/IP.

Tip: The host name value is typically specified on the TCPIP.DATA HOSTNAME statement.

Restartable subsystems

Configuring VMCF and TNF as restartable subsystems has the following advantages:

- Error detection is provided when the subsystems do not seem to be initializing properly.
- You can change the system name on the restart.
- Commands are available to remove users from internal tables, display current users and to terminate the subsystem.

In summary, a restartable VMCF and TNF configuration provides better availability and is therefore recommended.

If you choose to use restartable VMCF and TNF, follow these steps:

1. Update your IEFSSN*xx* member in SYS1.PARMLIB with the TNF and VMCF subsystem statements required by TCP/IP. The specification can be in either the IBM recommended keyword parameter form or the positional parameter form of IEFSSNxx. For example:

   ```
   * The keyword parameter form is:

   SUBSYS SUBNAME(TNF)
   SUBSYS SUBNAME(VMCF)

   * The positional parameter form is:
   TNF
   VMCF
   ```

2. Add procedure EZAZSSI to your system PROCLIB. A sample of this procedure is located in the SEZAINST library.

   ```
   //EZAZSSI PROC P=&SYSNAME.
   //STARTVT EXEC PGM=EZAZSSI,PARM=&P,TIME=1440
   ```

3. Start VMCF and TNF using the procedure EZAZSSI before starting TCP/IP. If your node name is the same as the MVS system symbolic &SYSNAME, then you can start VMCF and TNF with the following command:

   ```
   S EZAZSSI
   ```

   If your node name is different than the MVS system symbolic &SYSNAME, start VMCF and TNF with the following command:

   ```
   S EZAZSSI,P=nodename
   ```

   Replace *nodename* with the system name of your MVS system.

Non-restartable subsystems

If you will not be using restartable VMCF and TNF, you should update your IEFSSN*xx* member in SYS1.PARMLIB with the following subsystem statements required by TCP/IP. The specification can be in either the IBM suggested keyword parameter form or the positional parameter form of IEFSSN*xx*.

Restriction: If you are using the keyword parameter form and your IEFSSN*xx* member contains the BEGINPARALLEL statement, the TNF and VMCF statements must precede the BEGINPARALLEL statement, so that TNF and VMCF are initialized serially instead of in parallel. VMCF requires that TNF is initialized first.

The keyword parameter form is:

```
SUBSYS SUBNAME(TNF) INITRTN(MVPTSSI)
SUBSYS SUBNAME(VMCF) INITRTN(MVPXSSI) INITPARM(nodename)
```

The positional parameter form is:

```
TNF,MVPTSSI
VMCF,MVPXSSI,nodename
```

Replace *nodename* on the VMCF line with the system name of your MVS system.

VMCF commands

If you will be using restartable VMCF, the following VMCF commands let you display the names of the current users of VMCF and TNF, and if necessary, remove names from the name lists.

Note: Removing names from the name lists and stopping either subsystem can have unwanted results, if done hastily. Use the REMOVE and stop (P) commands carefully and only as a last resort.

If you remove a user, the application is not canceled, nor is the connection severed. In other words, the removed application may remain active in the system, and may subsequently abend 0D6/0D4/0C4, or cause TCP/IP to hang. A user that is removed from VMCF may still be a user of TNF and even TCP/IP, and vice versa.

To terminate users and stop VMCF or TNF properly, follow these steps:

1. Display the current users of the subsystems by using one of the following commands:

   ```
   F VMCF,DISPLAY,NAME=*
   ```

   ```
   F TNF,DISPLAY,NAME=*
   ```

2. Terminate those users. If termination fails, use the REMOVE command as a last resort to force them from the name list.
3. Stop the subsystem, using one of the following commands:

   ```
   P VMCF
   ```

```
P TNF
```

If the P command fails, use one of the following commands:

```
FORCE ARM VMCF
```

```
FORCE ARM TNF
```

The following list describes the commands:

F TNF,DISPLAY,NAME=[name|*]
Displays the named user [or all (*) users] of TNF, sorted by ASID.

F TNF,REMOVE,NAME=[name|*]
Removes either the named user [or all (*) users] from the TNF internal tables.

P TNF
Requests TNF to terminate.

F VMCF,DISPLAY,NAME=[name|*]
Displays the named user [or all (*) users] of VMCF, sorted by name.

F VMCF,REMOVE,NAME=[name|*]
Removes either the named user [or all (*) users] from the VMCF internal tables.

P VMCF
Requests VMCF to terminate

Sample commands:

```
F TNF,DISPLAY,NAME=TCPV3
F VMCF,DISPLAY,NAME=*
F TNF,REMOVE,NAME=FTPSERV
F VMCF,REMOVE,NAME=*
P TNF
```

Common VMCF problems

The following problems are some common VMCF problems:

- VMCF or TNF fail to initialize with an 0C4 abend.

  This is probably an installation problem; check the program properties table (PPT) entries for errors. Some levels of MVS do not flag PPT syntax errors properly.

- Abends 0D5 and 0D6 after REMOVEing a user.

  This is probably because the application is still running and using VMCF. It is not recommended that users be removed from VMCF or TNF without first terminating the affected user.

- VMCF or TNF do not respond to commands.

  This is probably because one or both of the non-restartable versions of VMCF or TNF are still active. To get them to respond to commands, stop all VMCF/TNF users, FORCE ARM VMCF and TNF, then use EZAZSSI to restart.

- VMCF or TNF cannot be stopped.

  This is probably because users still exist in the VMCF and TNF lists. Use the F VMCF,DISPLAY,NAME=* and F TNF,DISPLAY,NAME=* commands to identify those users who are still active. Then either cancel those users or remove them from the lists using the F VMCF,REMOVE and F TNF,REMOVE commands.

- Address Space Identifiers (ASIDs) become nonreusable when VMCF and TNF address spaces are stopped or restarted.

  Because VMCF and TNF address spaces provide PC-entered services that must be accessible to all address spaces, they each obtain a system LX. This causes the ASIDs associated with these address spaces to be nonreusable when these address spaces are terminated. If VMCF and TNF are terminated enough times all available ASIDs could be exhausted, preventing the creation of new address spaces

on the system. In this case, an IPL will be required. For more information on tuning parameters for the maximum number of ASIDs in a system, see the MAXUSER parameter in z/OS MVS Initialization and Tuning Reference.

IUCV/VMCF considerations

The IUCV/VMCF inter-address space communication API enables applications running in the same MVS image to communicate with each other without requiring the services of the TCP/IP protocol stack. The VMCF/TNF subsystems provide these services, which are still available in z/OS Communications Server. Several components of TCP/IP in z/OS Communications Server continue to make some use of these services for the purpose of inter-address space communications. These include:

- The AF_IUCV domain sockets for the TCP/IP C socket interface. The AF_IUCV domain enables applications executing in the same z/OS image and using the TCP/IP C socket interface to communicate with each other using a socket API, but without requiring the services of the TCP/IP protocol stack, as no network flows result in these communications. This is quite different from the more common AF_INET domain that enables socket communication over an IP network. AF_IUCV sockets continue to be supported in z/OS Communications Server.

  An example of a TCP/IP-provided application that exploits AF_IUCV sockets is the SNMP Query Engine component (SQESERVE). The z/OS UNIX socket library provides a similar functionality to the AF_IUCV domain sockets with its AF_UNIX domain. Users creating new applications should consider using AF_UNIX domain sockets.
- The Pascal socket interface also makes use of the IUCV/VMCF services for a limited set of inter-address space communication flows. As a result, any applications (provided by IBM or others) that use the Pascal socket API also still have a requirement for the VMCF/TNF subsystems. TCP/IP provides several applications and commands that use these interfaces, such as the LPD servers, and the TSO TELNET, HOMETEST, TESTSITE, and LPR commands. IUCV/VMCF services require the usage of an address space name of SYSTEM. This means a TSO user cannot have the user ID name of SYSTEM.

Therefore, in z/OS Communications Server you must continue to configure and start the VMCF and TNF subsystems as you did in TCP/IP V3R2. However, because the VMCF/TNF subsystems are no longer used to communicate directly with the TCP/IP protocol stack in z/OS Communications Server, the amount of CPU they will consume will be significantly lower than in the TCP/IP V3R2 environment.

Step 4: Update the VTAM application definitions

You must update the VTAM definitions for the TN3270E Telnet server. See “Steps for customizing the VTAM configuration data set for Telnet” on page 573 for an example of the VTAM definitions.

SEZAINST(VTAMLST) contains a sample of the VTAM definitions for TN3270E Telnet server applications. You should copy this member, update it, and add it to the ATCCON*xx* member of VTAMLST. This ensures that the TN3270E Telnet server applications are activated when VTAM is started.

Because the TCP/IP LU code cannot handle multiple concurrent sessions, you must code SESSLIM=YES for each TN3270E Telnet server LU defined to VTAM. Otherwise, if SESSLIM=NO, menu or session manager applications that use return session processing might cause session termination.

Step 5: Verify that the required address spaces are active

TCP/IP uses services from other address spaces in its processing. While TCP/IP waits for availability of those services, it is recommended that you verify that the OMVS, resolver, and VTAM address spaces have completed initialization before starting TCP/IP. If using automation to start TCP/IP, have it look for the BPXI004I, EZZ9291I, and IST020I messages, respectively, before issuing the START command.

For information on how the resolver can be started, see “Starting the resolver” on page 737. You can use the resolver's MODIFY DISPLAY command to check that the resolver is active and what resolver setup statements are being used. For the syntax and usage of the command, see z/OS Communications Server: IP System Administrator's Commands.

Step 6: Start the TCP/IP address space

Enter the MVS START command from the operator's console to start TCP/IP, specifying the member name of your cataloged procedure. This will start the TCP/IP address space and any of the servers you have defined in the AUTOLOG statement in PROFILE.TCPIP. For example, if the procedure to start the TCP/IP address space was called TCP1 in your PROCLIB, you would enter:

```
START TCP1
```

For information on updating the TCPIP cataloged procedure or configuration statements used to configure the TCPIP address space, see z/OS Communications Server: IP Configuration Reference.

Step 7: Set up cataloged procedures and configuration data sets

At this point in the configuration process, you can choose to either set up procedures or you can do each one individually when you set up the appropriate application, function, or server.

See the appropriate topics for more information about setting up a particular application, function, or server.

Chapter 3. Security

The z/OS Communications Server, along with other elements of z/OS, provide numerous enterprise-strength security services to protect your mission-critical data. This topic provides an overview of these technologies and how they can be used for a safe and secure z/OS TCP/IP deployment.

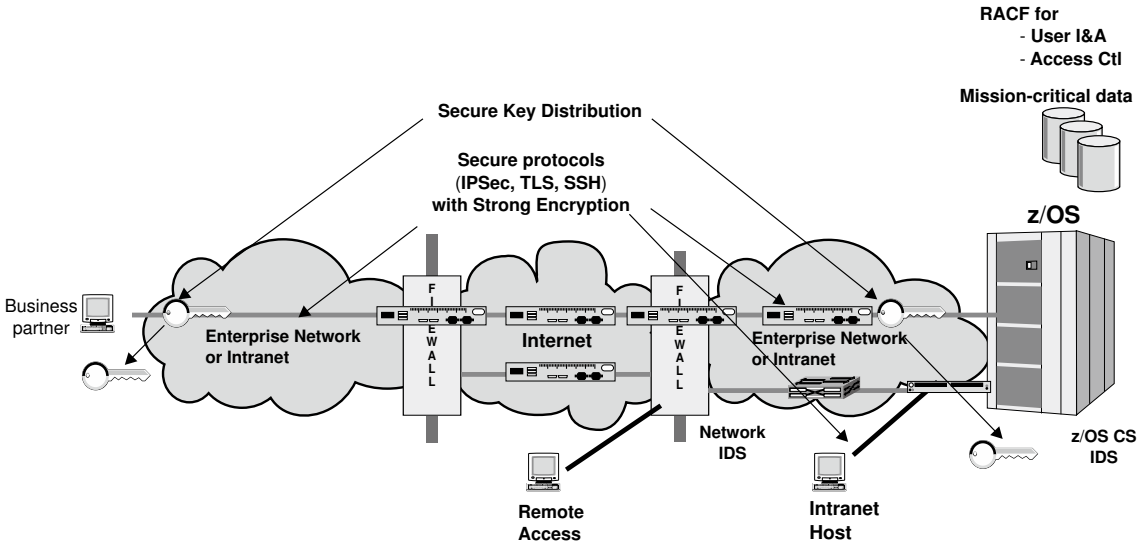


Figure 26. Elements of a secure TCP/IP deployment

Tip: Many of the tasks, examples, and references in this information assume that you are using the z/OS Security Server (RACF). References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on task performance.

The Communications Server protects data and other resources on the system. Communications Server applications use RACF services to ensure that users requesting application access are identified and authenticated, and to protect data and other system resources from unauthorized access. The Communications Server safeguards the availability of the system by protecting against denial of service attacks from the network.

The Communications Server protects data in the network by supporting a variety of cryptographic-based network security protocols such as IPSec and TLS/SSL. z/OS also supports the SSH protocol. These security protocols ensure the identity of the communication partners (partner authentication), that data received is originated by the claimed sender (data origin authentication), that contents were unchanged in transit (message integrity), and that sensitive data is concealed using encryption (data privacy).

The Communications Server provides security event reporting to record potential security violations. These services may help you identify potential sources of subsequent attacks, respond more quickly to network attacks, and manage system resources during periods of high network traffic for key applications.

Note: Some of the security features described in this information have not yet been implemented for IPv6. To determine which functions are supported for IPv6, see the IPv6 support tables in z/OS Communications Server: IPv6 Network and Application Design Guide.

Application security

The Communications Server protects data and other system resources accessed by applications included in the Communications Server element. This protection requires verification of the identity of the user requesting access. This process is called identification and authentication. In addition, access to resources must be limited to those users with permission. This process is called access control.

© Copyright IBM Corp. 2000, 2022

Communications Server applications use RACF for identification and authentication, and access control decisions. Authenticated users are granted access to RACF resources only for which they have permission

Some applications allow anonymous access. Applications that allow anonymous access include anonymous FTP and Remote Execution. The Communications Server ensures that all anonymous access can be controlled by the installation. If anonymous access is allowed, the resources accessed can be limited in several ways:

- The application can be configured to limit resources for which access will be attempted.
- The application can be configured to use a RACF user ID to represent the anonymous user. In this case, access is allowed for those resources specifically permitted for the anonymous RACF user ID and for those resources that are universally accessible.

Most Communications Server applications must be configured specifically to allow anonymous access.

Table 9 on page 124 depicts a representative set of Communications Server applications, whether user identification is required, and the security credentials under which resource access is made. For more information on specific application considerations, see the topic about each application.

Table 9. User identification, authentication, and access control for z/OS Communications Server applications

Server	End-user identification	Resource access
FTP	Optional [1]	End-user ID or configured anonymous user ID [2]
LPD	Optional [1]	Server ID or end-user ID
MVS REXECD	Required	End-user ID
MVS RSHD	Required (password optional) [1]	Surrogate user ID or end-user ID
NSSD [network security services (NSS) server]	Required	NSS client user ID
Policy Agent server	Required	Policy client user ID
UNIX REXECD	Required	End-user ID
UNIX RSHD	Required (password optional) [1]	End-user ID or Server user ID (exit routine to verify request)
UNIX shell (Telnet/rlogin)	Required	End-user ID

1. All items listed as optional are installation controlled and can all be configured to require full end-user identification.
2. Accessible files can be configured on a server basis to limit access.

TCP/IP resource protection

The Communications Server uses the System Authorization Facility (SAF) to protect TCP/IP resources from unauthorized access. These resources are represented by resource profiles defined in the SERVAUTH class. When describing resource profile names, the following conventions are used:

- *sysname* is the MVS system name. Substitute your system name.
- *tcpname* is the TCP/IP job name. Substitute your job name.
- *ftpdaemonname* is the job name of the FTP daemon. Substitute your FTPD job name. If your FTPD job name contains fewer than 8 characters, substitute the job name of the process that is started by FTPD, which is usually the original job name with the number 1 appended.

- *rpcbindname* is the job name of rpcbind. Substitute your rpcbind job name. If your rpcbind job name contains fewer than 8 characters, substitute the job name of the process that is started by rpcbind, which is usually the original job name with the number 1 appended.

The use of SERVAUTH is optional. The installation can choose to use any combination of the protections provided by SERVAUTH.

In addition to the use of SERVAUTH protection, further resource protection is provided through such functions as intrusion detection services (IDS), syslogd isolation, and IP filtering, or through controlling access to the VARY TCPIP command.

Local user access control to TCP/IP resources using SAF

You can use System Authorization Facility (SAF) to control which z/OS users can access specific TCP/IP resources, which protects against unauthorized user access to these resources.

You define SAF resource profiles in the SERVAUTH class to control access to the TCP/IP resources. After you define a SAF resource profile, a local user can access the associated TCP/IP resource if their user ID has at least READ access to the resource.

z/OS Communication Server programs call SAF to determine which users have access to protected resources. The user's credentials, a resource name, and a requested level of access (READ, UPDATE, and so on) are provided to SAF. SAF has three defined return codes:

0
Permit

4
No decision

8
Deny

The following situations can result in a no-decision return code from SAF:

- The security server is not available.
- The resource class is not active.
- The named resource does not have a profile defined in the class.

When SAF returns a no-decision return code, the resource manager decides whether to allow access. The No SAF decision column in Table 10 on page 125 indicates the action that the resource manager takes for each resource.

Table 10 on page 125 summarizes the SERVAUTH resource names that are used by TCP/IP.

Table 10. SERVAUTH resource names used by TCP/IP

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
Broadcast access control	Provides ability to control whether an application is permitted to set the SO_BROADCAST socket option needed to send broadcast datagrams	Permit	EZB.SOCKOPT.*sysname*.*tcpname*.SO_BROADCAST
			TCPIP SOCKOPT ACCESS CHECK

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
CIM provider access control	Provides ability to restrict access to CIM data	Deny	EZB.CIMPROV.*sysname.tcpname* TCPIP CIM PROVIDER CHECK
DCAS server access control	Controls ability to access DCAS server based on SAF user ID associated with TLS-authenticated X.509 client certificate	Permit	EZA.DCAS.*cvtsysname* DCAS SAFCERT CHECK FOR USER certuser or TCPIP EZACDRAU AUTH CHECK FOR EZA.DCAS.*cvtsysname*
Fast Response Cache Accelerator (FRCA) Access Control	Provides ability of user to create FRCA cache (FRCA used by web servers for caching static web pages in the stack)	Deny, see result 1	EZB.FRCAACCESS.*sysname.tcpname* TCPIP FRCA ACCESS CHECK
FTP server access control	Controls ability to access FTP server based on SAF user ID used to log in	Permit	EZB.FTP.*sysname.ftpdaemonname*.PORT*xx xxx* (none)
FTP SITE command control	Provides ability to restrict usage of SITE DUMP and DEBUG commands (commands generate large amount of output)	Permit	EZB.FTP.*sysname.ftpdaemonname*.SITE.DUMP EZB.FTP.*sysname.ftpdaemonname*.SITE.DEBUG (none)
FTP z/OS UNIX file system access control	Provides ability to generally restrict FTP user access to the z/OS UNIX file system	Permit	EZB.FTP.*sysname.ftpdaemonname*.ACCESS.HFS (none)
ipsec command access control	Provides ability to control **ipsec** command usage	Deny	EZB.IPSECCMD.*sysname.tcpname.command_type* EZB.IPSECCMD.*sysname*.DMD_GLOBAL.*command_type* TCPIP EZACDRAU AUTH CHECK FOR EZB.IPSECCMD.*sysname.tcpname.command_type* or TCPIP EZACDRAU AUTH CHECK FOR EZB.IPSECCMD.*sysname*.DMD_GLOBAL.*command_type*

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
IPSec network management interface (NMI) access control for control requests (local)	Controls whether a user can issue NMI control requests to the local IKE daemon to manage IP filtering and IPSec function (for example, activate and deactivate requests) pertaining to a local TCP/IP stack	Deny	EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.CONTROL TCPIP EZACDRAU AUTH CHECK FOR EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.CONTROL
IPSec NMI access control for display requests (local)	Controls whether a user can issue NMI monitoring requests to the local IKE daemon to retrieve IP filtering and IPSec monitoring data pertaining to a local TCP/IP stack	Deny	EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.DISPLAY TCPIP EZACDRAU AUTH CHECK FOR EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.DISPLAY
IPSec NMI and **ipsec** command access control	Controls whether a user can issue: • NMI requests to display IKE daemon NSS client information • The **ipsec** command with the `-w` option to display IKE daemon NSS IPSec client information	Deny	EZB.NETMGMT.*sysname*.*sysname*.IKED.DISPLAY TCPIP EZACDRAU AUTH CHECK FOR EZB.NETMGMT.*sysname*.*sysname*.IKED.DISPLAY

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
IPSec NMI and **ipsec** command access control for control requests (remote)	Controls whether a user can issue: • NMI management requests to the NSS server that pertain to an NSS client (for example, activate and deactivate requests) • The **ipsec** command with the -z option to perform a management action to an NSS IPSec client (for example, to activate and deactivate options)	Deny	EZB.NETMGMT.*sysname*.*clientname*.IPSEC.CONTROL TCPIP EZACDRAU AUTH CHECK FOR EZB.NETMGMT.*sysname*.*clientname*.IPSEC.CONTROL
IPSec NMI and **ipsec** command access control for display requests (remote)	Controls whether a user can issue: • NMI monitoring requests to the NSS server that pertain to an NSS client (that is, get requests) • The **ipsec** command with the -z option to display options for an NSS IPSec client	Deny	EZB.NETMGMT.*sysname*.*clientname*.IPSEC.DISPLAY TCPIP EZACDRAU AUTH CHECK FOR EZB.NETMGMT.*sysname*.*clientname*.IPSEC.DISPLAY

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
IPv6 Advanced Socket API access control	Provides ability to control whether an application is permitted to set IPv6 advanced socket API options: IPv6_NEXTHOP IPv6_TCLASS IPv6_RTHDR IPV6_HOPOPTS IPV6_DSPOPTS IPV6_RTHDRDSTOPT IPV6_PKTINFO IPV6_HOPLIMIT	Deny, see result 2	EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_NEXTHOP EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_TCLASS EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_RTHDR EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPOPTS EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_DSTOPTS EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_RTHDRDSTOPTS EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_PKTINFO EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPLIMIT
			TCPIP SOCKOPT ACCESS CHECK
Netstat command access control	Provides ability to restrict Netstat usage	Permit, see result 3	EZB.NETSTAT.*sysname*.*tcpname*.*netstat_option*
			TCPIP EZACDNET AUTH CHECK FOR EZB.NETSTAT.*sysname*.*tcpname*.*netstat_option*
Network security services (NSS) NMI and command access control	Controls whether a user can issue: • NMI requests to display connections to the NSS server • The **ipsec** command with the -x option to display NSS IPSec client connections to the NSS server • The **nssctl** command to display NSS client connections to the NSS server.	Deny	EZB.NETMGMT.*sysname*.*sysname*.NSS.DISPLAY
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NETMGMT.*sysname*.*sysname*.NSS.DISPLAY

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
NSS server access control	Controls whether an NSS IPSec client can register with the NSS server for the NSS IPSec certificate service	Deny	EZB.NSS.*sysname*.*clientname*.IPSEC.CERT
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSS.*sysname*.*clientname*.IPSEC.CERT
NSS server access control	Controls whether an NSS IPSec client can register with the NSS server for the NSS IPSec remote management service	Deny	EZB.NSS.*sysname*.*clientname*.IPSEC.NETMGMT
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSS.*sysname*.*clientname*.IPSEC.NETMGMT
NSS server access control	Controls whether an NSS XMLAppliance client can register with the NSS server for the XMLAppliance SAFAccess service.	Deny	EZB.NSS.*sysname*.*clientname*.XMLAPPLIANCE.SAFACCESS
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSS.*sysname*.*clientname*.XMLAPPLIANCE.SAFACCESS
NSS server access control	Controls whether an NSS XMLAppliance client can register with the NSS server for the XMLAppliance certificate service.	Deny	EZB.NSS.*sysname*.*clientname*.XMLAPPLIANCE.CERT
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSS.*sysname*.*clientname*.XMLAPPLIANCE.CERT
NSS server access control	Controls whether an NSS XMLAppliance client can register with the NSS server for the XMLAppliance private key service.	Deny	EZB.NSS.*sysname*.*clientname*.XMLAPPLIANCE.PRIVKEY
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSS.*sysname*.*clientname*.XMLAPPLIANCE.PRIVKEY
NSS server certificate access control	Controls whether an NSS client can access a CERTAUTH certificate on the key ring of the NSS server	Deny	EZB.NSSCERT.*sysname*.*mappedlabelname*.CERTAUTH
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSSCERT.*sysname*.*mappedlabelname*.CERTAUTH

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
NSS server certificate access control	Controls whether an NSS client can access a PERSONAL or SITE certificate on the key ring of the NSS server	Deny	EZB.NSSCERT.*sysname*.*mappedlabelname*.HOST
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSSCERT.*sysname*.*mappedlabelname*.HOST
NSS server private key access control	Controls whether an NSS XMLAppliance client can access the private key for a certificate on the key ring of the NSS server	Deny	EZB.NSSCERT.*sysname*.*mappedlabelname*.PRIVKEY
			TCPIP EZACDRAU AUTH CHECK FOR EZB.NSSCERT.*sysname*.*mappedlabelname*.PRIVKEY
OSM access control	Controls ability to access the intranode management network using OSM interfaces	Deny	EZB.OSM.*sysname*.*tcpname*
			TCPIP OSM ACCESS CHECK
Partner information ioctl access control	Controls whether an application can use the SIOCGPARTNERINFO ioctl to obtain partner security credentials within a sysplex or subplex over a trusted TCP connection	Deny	EZB.IOCTL.*sysname*.*tcpprocname*.PARTNERINFO
			SIOCGPARTNERINFO
Policy Agent command control	Provides ability to restrict **pasearch** command, IKE daemon, policy clients, and nslapm2 usage by type	Deny	EZB.PAGENT.*sysname*.*image*.*ptype*
			TCPIP EZACDRAU AUTH CHECK FOR EZB.PAGENT.*sysname*.*image*.*ptype*
Real-time application-controlled TCP/IP trace NMI access control - Open request	Controls whether an application can invoke the NMI to open a trace; intended for network management applications	Deny	EZB.TRCCTL.*sysname*.*tcpname*.OPEN
			TCPIP NETWORK MANAGEMENT

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
Real-time application-controlled TCP/IP trace NMI access control - Set filters	Controls whether an application can invoke the NMI to set filters for packet trace; intended for network management applications	Deny	EZB.TRCCTL.*sysname*.*tcpname*.PKTTRACE TCPIP NETWORK MANAGEMENT
Real-time application-controlled TCP/IP trace NMI access control - Set filters	Controls whether an application can request IPSec cleartext data on a packet trace filter	Deny	EZB.TRCSEC.*sysname*.*tcpname*.IPSEC TCPIP NETWORK MANAGEMENT
Real-time application-controlled TCP/IP trace NMI access control - Set filters	Controls whether an application can invoke the NMI to set filters for data trace; intended for network management applications	Deny	EZB.TRCCTL.*sysname*.*tcpname*.DATTRACE TCPIP NETWORK MANAGEMENT
Real-time application-controlled TCP/IP trace NMI access control - Set filters	Controls whether an application can request AT-TLS cleartext data on a data trace filter	Deny	EZB.TRCSEC.*sysname*.*tcpname*.ATTLS TCPIP NETWORK MANAGEMENT
Real-time OSAENTA information service access control	Provides ability to restrict access to select real-time OSAENTA packet trace records accessible using the OSAENTA information service; intended for network management applications	Deny, see result 4	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPOT TCPIP NETWORK MANAGEMENT

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
Real-time SMF information service access control	Provides ability to restrict access to select real-time SMF records accessible using the SMF information service; intended for network management applications	Deny, see result 4	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPSM TCPIP NETWORK MANAGEMENT
Real-time TCP connection information service access control	Provides ability to restrict access to the TCP connection information using TCP connection information service; intended for network management applications	Deny, see result 4	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPCN TCPIP NETWORK MANAGEMENT
Real-time TCP/IP packet trace service access control	Provides ability to restrict access to select real-time packet trace records accessible using the TCP/IP packet trace service; intended for network management applications	Deny, see result 4	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPDA TCPIP NETWORK MANAGEMENT
Real-time zERT connection detail service access control	Provides ability to restrict access to z/OS Encryption Readiness Technology information using the zERT connection detail service; intended for network management applications	Deny, see result 4	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPER TCPIP NETWORK MANAGEMENT

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
Real-time zERT summary service access control	Provides ability to restrict access to z/OS Encryption Readiness Technology information using the zERT summary service; intended for network management applications	Deny, see result 4	EZB.NETMGMT.*sysname*.*tcpname*.SYSTCPES TCPIP NETWORK MANAGEMENT
rpcbind access control	Provides ability to control whether an applications is permitted to register and unregister its port with rpcbind.	Deny	EZB.RPCBIND.*sysname*.*rpcbindname*.REGISTRY (none)
SNMP agent control	Provides control over usage of SNMP subagents that connect to the SNMP agent by using a TCP connection	Permit	EZB.SNMPAGENT.*sysname*.*tcpname* TCPIP EZACDRAU AUTH CHECK FOR EZB.SNMPAGENT.*sysname*.*tcpname*
TCP/IP local port access control	Controls user ability to bind to a non-ephemeral TCP or UDP port	Deny	EZB.PORTACCESS.*sysname*.*tcpname*.*port_safname* TCPIP PORT ACCESS CHECK PORT *portnum*
TCP/IP netaccess access control	Controls local user inbound and outbound access to network resources, and local user access to local IP address when explicitly binding to local interface (or using job-specific or destination-specific source IP addresses)	Deny	EZB.NETACCESS.*sysname*.*tcpname*.*zonename* TCPIP NETWORK ACCESS CHECK *ipaddress*
TCP/IP stack access control	Controls user ability to open a socket and get host name or host ID	Permit	EZB.STACKACCESS.*sysname*.*tcpname* TCPIP STACK ACCESS CHECK

Table 10. SERVAUTH resource names used by TCP/IP (continued)

Function	Description	No SAF decision	SERVAUTH resource name LOGSTR in any SAF logging (SMF type 80 records for RACF)
TCP/IP stack initialization access control	Controls ability of applications to open a socket before AT-TLS policy is loaded into the TCP/IP stack	Deny	EZB.INITSTACK.*sysname.tcpname* TCPIP INIT STACK ACCESS CHECK
TN3270E Telnet server access control	Controls ability to access TN3270E Telnet server based on SAF user ID associated with TLS-authenticated X.509 client certificate	Deny	EZB.TN3270.*sysname.tn3270name*.PORT*x xxxx* TN3270 SAFCERT CHECK FOR USER userid PORT *portnum* ON *tn3270name*
VIPARANGE access control for any VIPA range (bind)	Controls whether an application can create a DVIPA by binding to a DVIPA that is specified by any VIPARANGE statement	Permit	EZB.BINDDVIPARANGE.*sysname.tcpname* TCPIP BINDDVIPA ACCESS CHECK
VIPARANGE access control for any VIPA range (MODDVIPA and ioctl)	Provides access control for all VIPARANGE statements, and controls whether a user or application can perform the following tasks: • Create a dynamic VIPA (DVIPA) that is specified by any VIPARANGE statement, using the SIOCSVIPA ioctl call, the SIOCSVIPA6 ioctl call, or the MODDVIPA utility • Delete a DVIPA that was created using this profile and the SIOCSVIPA ioctl call, the SIOCSVIPA6 ioctl call, or the MODDVIPA utility	Deny, see result 5	EZB.MODDVIPA.*sysname.tcpname* TCPIP MODDVIPA or SIOCSVIPA(6) ACCESS CHECK

Table 10. SERVAUTH resource names used by TCP/IP (continued)			
			SERVAUTH resource name
Function	**Description**	**No SAF decision**	**LOGSTR in any SAF logging (SMF type 80 records for RACF)**
VIPARANGE access control for a specific VIPA range (bind)	Controls whether an application can create an application-specific DVIPA, by binding to a DVIPA that is specified by a VIPARANGE statement that includes the SAF parameter with the same value for *resname*.	Deny	EZB.BINDDVIPARANGE.*sysname.tcpname.resname* TCPIP BINDDVIPA SAF ACCESS CHECK
VIPARANGE access control for a specific VIPA range (MODDVIPA and ioctl)	Provides access control for a specific VIPARANGE statement that includes the SAF parameter with the same value for *resname*, and controls whether a user or application can perform the following tasks: • Create an application-specific DVIPA that is specified by a specific VIPARANGE statement, using the SIOCSVIPA ioctl call, the SIOCSVIPA6 ioctl call, or the MODDVIPA utility • Delete a DVIPA that was created using this profile and the SIOCSVIPA ioctl call, the SIOCSVIPA6 ioctl call, or the MODDVIPA utility	Deny	EZB.MODDVIPA.*sysname.tcpname.resname* TCPIP MODDVIPA or SIOCSVIPA(6) SAF ACCESS CHECK

Table 10. SERVAUTH resource names used by TCP/IP (continued)			
			SERVAUTH resource name
Function	**Description**	**No SAF decision**	**LOGSTR in any SAF logging (SMF type 80 records for RACF)**
Results: 1. Deny, unless the user ID is a WLM user or is a UNIX System Services superuser. 2. Deny, unless the user ID is APF authorized or is a UNIX System Services superuser. 3. Permit, except for the DROP option, when access is denied. 4. Deny, unless the user ID is a UNIX System Services superuser or has READ access to BPX.SUPERUSER. 5. Deny, unless the user ID is APF authorized and is a UNIX System Services superuser.			

Stack access control

You can create a SAF resource profile to control access to a TCP/IP stack. There are no new TCP definitions required. The resource profile controls whether users or groups of users have access to the TCP/IP stack by controlling their ability to open an AF_INET or AF_INET6 socket and to obtain the host ID or host name. Create the EZB.STACKACCESS.*sysname.tcpname* resource profile in the SERVAUTH class for the TCP/IP stack to be protected. After you define this resource profile, permit users to the profile and grant them READ access to the resource. If a user does not have READ access to the resource for a stack, the user cannot access the stack. If you do not define a resource profile for a stack, all users have access to that stack.

Guideline: Some security products do not distinguish between a resource profile that is not defined and a user that is not permitted to that resource profile. If your product does not make this distinction, you must define the stack access resource profile and permit users to it whenever the SERVAUTH class is active.

Figure 27 on page 137 provides an overview of stack access control. *sysname* refers to the MVS system variable sysname. *tcpname* refers to the TCP/IP job name. User Tom has permission to access both Stack1 and Stack2, Joe does not have permission to access any stack, and Bob has permission to access Stack2 but not Stack1.

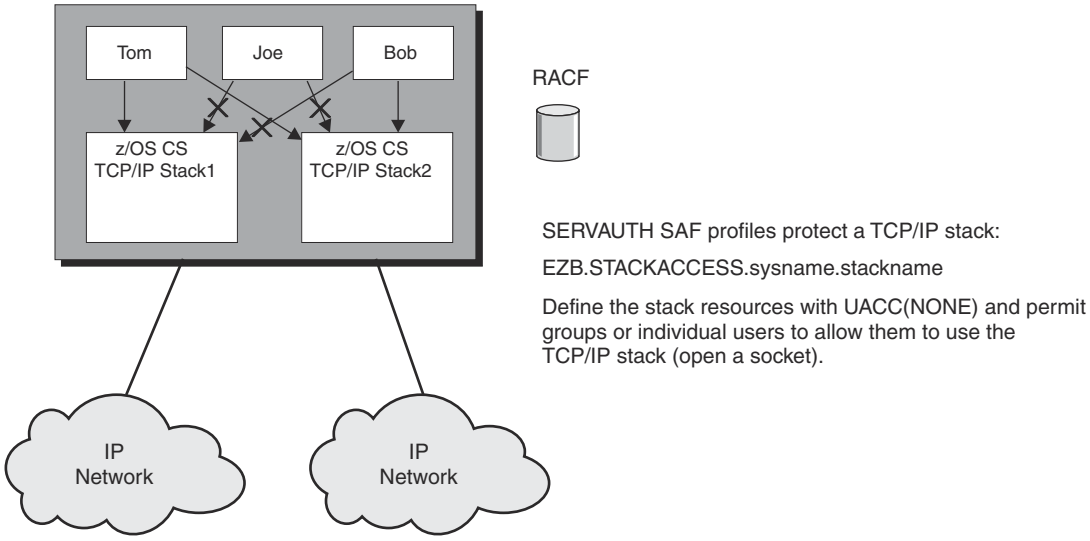


Figure 27. Stack access control overview

Port access control

You can use port access control to protect against unauthorized use of ports. You can control an application's ability to explicitly bind to, or listen on, specific TCP and UDP ports or port ranges by either reserving particular ports or by controlling access to unreserved ports.

Controlling access to particular ports

You can control access to particular ports by port number, by reserving the port using the PORT or PORTRANGE profile statements. Use the PORT and PORTRANGE statements to reserve well-known or configured ports for the applications that need to bind to them. You can use the optional SAF parameter to provide additional access control.

You can reserve an individual port or range of ports with a job name, a wildcard job name (*), a partial wildcard job name (1 - 8 characters with one or more wildcard characters), or the special job name of RESERVED. The asterisk wildcard character can be used to indicate zero or more unspecified characters. The question mark wildcard character can be used to indicate a single unspecified character. If you specify a job name, the port is reserved for an application that has the specified job name. If you specify a partial wildcard job name, the port is reserved for any application that matches the partial wildcard job name. If you specify a wildcard job name, the port is reserved for any application with any job name. The RESERVED job name shuts down the use of a port or range of ports for any application.

If you specify the SAF keyword on the PORT or PORTRANGE statement, it can provide additional access control by verifying that the user ID associated with an application at the time of a bind to the port is authorized to access the port. The SAF keyword value specifies a portion of the resource name that represents the port. Define the EZB.PORTACCESS.*sysname.stackname.port_safname* resource profile in the SERVAUTH class to control access to the port, where *port_safname* is the same value that you specify on the SAF keyword of the PORT or PORTRANGE statement. The user ID that is associated with the application at the time of the bind request must have READ access to this resource for the application to be able to bind to the port.

Figure 28 on page 138 provides an overview of port access control. In this example, z/OS user WEBSERV (web server) is permitted to bind to port 80. User Bob is not permitted to bind to port 80.

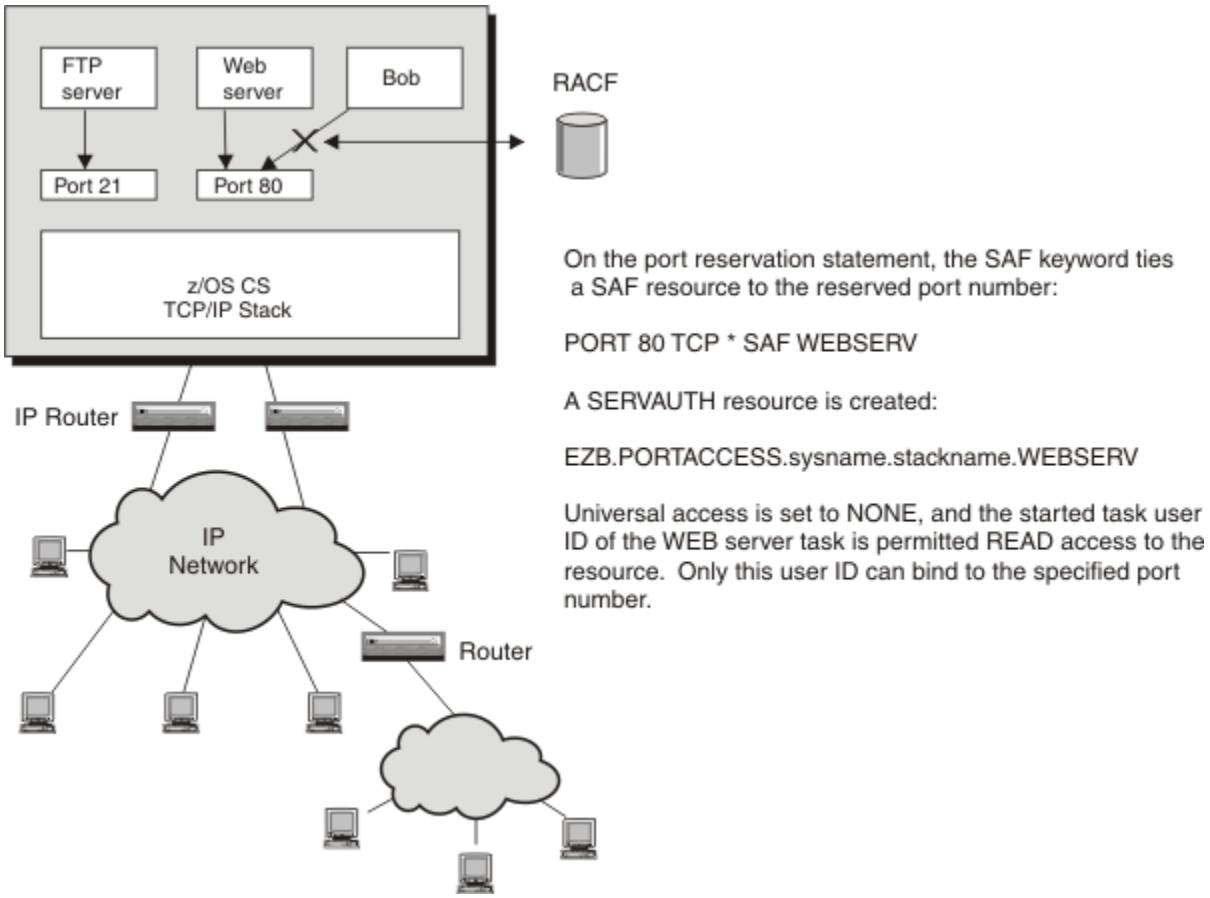


Figure 28. Port access control overview

Controlling access to unreserved ports

You can control which applications are allowed to access ports that have not been reserved by a PORT or PORTRANGE statement. You can use the PORT statement with the UNRSV keyword, or the RESTRICTLOWPORTS parameter on the UDPCONFIG or TCPCONFIG statements.

Using the PORT statement to control access to all unreserved ports

In addition to reserving particular ports, you can also use the PORT statement to control application access to user-specified ports that have not been reserved. Access control is applied when an application issues an explicit bind or a TCP listen.

You can use job names or SAF resources, or a combination of both, to control access to user-specified unreserved ports, or you can unconditionally deny access to user-specified unreserved ports. You do this by configuring one or more PORT statements in which the port number is replaced by the keyword UNRSV.

Result: PORT UNRSV statements control access to nonzero, unreserved ports specified on explicit binds. Access to unreserved ports that are assigned by the stack is not affected.

If you want to use PORT UNRSV access control, see the following considerations:

- Using the parameters on the PORT UNRSV statement

  To use only job names to control unreserved port access, configure one or more PORT UNRSV statements using the JOBNAME keyword with specific job names or partial wildcard job names (1 - 8 characters with one or more wildcard characters). If you specify the wildcard job name (*), the PORT UNRSV statement applies to all job names. If an application's job name matches more than one PORT UNRSV statement, the statement with the closest matching specified job name is used to control access to unreserved ports.

  To use a SAF resource instead of job names to control unreserved port access, configure a PORT UNRSV statement specifying the wildcard job name (*) and the SAF keyword and resource name. Because you can have only one PORT UNRSV statement with the wildcard job name per protocol, this method allows the use of only one SAF resource per protocol.

  To use a combination of job names and a SAF resource to control access, specify the SAF keyword and resource name on one or more PORT UNRSV statements with different job names. You can specify a different SAF resource on each of these PORT UNRSV statements.

  If you do not configure any PORT UNRSV statements for a protocol, all applications are allowed to use unreserved ports. This is the default behavior. To change this behavior and unconditionally deny access to user-specified unreserved ports, configure a PORT UNRSV statement, specifying the wildcard job name (*) and the keyword DENY.

  For the UDP protocol, access is always controlled when an explicit bind is issued (WHENBIND). For UDP applications, there is no distinction between client or server roles because there is no explicit socket API that indicates that the application is a UDP server. All UDP applications that bind their socket to a specific local unreserved port do need to pass any configured UDP port access control checks.

  For the TCP protocol, you can control access to user-specified unreserved ports when the explicit bind is issued (WHENBIND) or when a listen is issued on that port (WHENLISTEN). WHENLISTEN access control is targeted to TCP applications acting as servers (that is, applications able to accept incoming client TCP connections) and is the default for TCP. If the default is used or WHENLISTEN is specified, and no listen is issued for the unreserved port, no access control check is made. WHENBIND access control can affect TCP client applications that bind to a specific local port for outbound TCP connections. Every PORT UNRSV statement for the TCP protocol must specify the same access control. You cannot specify WHENLISTEN on some statements and WHENBIND on other statements. If you do, the conflicting configuration statement will fail.
- Implementing PORT UNRSV access control in stages

  Using PORT UNRSV access control can have unexpected consequences. Consider implementing this function in the following stages:

  1. Determine the necessary port reservations statements for your applications.

For example, you might begin with 'PORT UNRSV TCP * SAF *xyz* WHENBIND', where the SERVAUTH profile has UACC(READ) and auditing of successes is enabled. This would give you an indication of how many applications currently bind to unreserved ports. Verify the applications reported, and if deemed valid, reserve their ports using the PORT or PORTRANGE statements.

2. Enable access control enforcement by specifying DENY or a SAF resource with UACC(NONE) and monitor failures.

   As failures are identified, configure appropriate PORT reservation statements to allow authorized application access to those ports.

- Enforcing PORT UNRSV access control

  If you configure one or more PORT UNRSV statements for a protocol, access is unconditionally denied to any application that explicitly binds to an unreserved port and does not match the protocol and job name on any of the configured PORT UNRSV statements. Applications that explicitly bind to an unreserved port and that do match the protocol and job name on a PORT UNRSV statement are allowed to access the unreserved port, unless the access is restricted by the SAF or DENY keywords. If the SAF keyword is specified, the user ID associated with the application that attempts to access the port must be permitted to the specified SAF resource. If the DENY keyword is specified, access is unconditionally denied.

  Rules: If you configure PORT UNRSV statements for UDP or for TCP with the WHENBIND option, the following rules apply:

  – Every application using that protocol that explicitly binds to a user-specified local unreserved port is subject to an authorization check. If the application does not have authority to access an unreserved port, the bind will fail.
  – For TCP, every client connection that first performs an explicit bind and explicitly specifies a local unreserved port is subject to an authorization check for access to unreserved ports.
  – Because client programs might run under many different user IDs, all address spaces in which the client program can run must be authorized (by job name, SAF resource, or both) to access an unreserved port.

    Alternatively, you can use the WHEN(PROGRAM) type of RDEFINE statements to authorize specific programs to a particular port, regardless of the address space in which they run. However, in this case these programs must be program-controlled resources.

  If you do not configure a PORT UNRSV statement for a protocol, then access to unreserved ports using that protocol is not controlled. If you do configure PORT UNRSV statements for a protocol, access is determined by the PORT UNRSV statement with the job name that most closely matches the application's job name; if the applications's job name does not match any of the PORT UNRSV statements, then access to unreserved ports is denied for that protocol.

Using the RESTRICTLOWPORTS parameter to control access to unreserved ports below port 1024

When the RESTRICTLOWPORTS parameter is specified on the UDPCONFIG or TCPCONFIG profile statements, an application cannot obtain a port in the range 1 - 1023 that has not been reserved by a PORT or PORTRANGE statement, unless the application is APF-authorized or has OMVS superuser [UID(0)] authority. z/OS Communications Server client applications that need to bind to a low port are provided as APF-authorized.

Tip: When you configure the RESTRICTLOWPORTS parameter on the TCPCONFIG or UDPCONFIG profile statements, PORT UNRSV statements for the corresponding protocol control access only to unreserved ports above port 1023.

Network access control

Network access control gives system administrators the ability to assign permission for z/OS users to access certain networks and hosts. With this function, the ability of users to send or receive data between z/OS and certain networks can be controlled through z/OS. Network access control provides an additional

layer of security to any authentication and authorization security that is used in the network or at the peer system by disallowing the unauthorized user to communicate with the peer network resource.

Essential elements of this function are as follows:

- The IP network is considered the resource to be protected.
- Use of the IBM zEnterprise® System (zEnterprise) intranode management network is protected by OSM access control and is exempt from network access control. For more information, see "OSM access control" on page 142.
- IP addresses are classified into *security zones*, in which each zone has a certain level of security sensitivity. A default security zone exists for interfaces that are not explicitly associated with a specific security zone. Security zones consist of one or more, perhaps discontiguous, IP address ranges that have the same security sensitivity and are identified by a specific zone name.
- SAF is used to check whether users or groups of users have READ access to a security zone.
- You define a SAF resource profile for each security zone and provide READ access to these resources to the users or groups of users that you want to have access to particular security zones. A security zone is represented by the EZB.NETACCESS.*sysname*.*tcpname*.*zonename* resource name in the SERVAUTH class.
- TCP/IP keeps a mapping of network resources by IP address to security zones. This mapping is consulted on certain inbound and outbound operations to determine the corresponding resource zone name for the most specific network defined. Then the current user's access to that resource is queried using SAF, and the operation is allowed or denied completion accordingly. This mapping is also consulted when the security ioctl is issued to extract the port of entry zone name of a socket's current peer.
- Network access control is used to control z/OS user access to a peer address in an IP network through a sockets application. Resource access checks occur at connection setup or acceptance time for TCP, peer identification time for UDP and RAW, and on the first and potentially subsequent sends or receives (TCP, UDP, or RAW) to a particular destination in a socket's lifetime.
- Network access control is used to control z/OS user access to local addresses when a socket is bound to a local address. Resource access checks occur when an application explicitly binds a socket to a local address, including the IPv4 address INADDR_ANY (0.0.0.0/32) or the IPv6 unspecified address, in6addr_any (::/128). Job-specific or destination-specific source IP addresses (designated by the SRCIP profile statement) are handled as if the application did an explicit bind to the configured address.
- Network access control security checks are made at the transport layer (TCP, UDP, and RAW). Other IP-specific packets generated by the stack are not covered under this function (such as ICMP echo replies, for example). Additionally, there is no user concept when dealing with packets that are being forwarded through the stack, and hence no checks are made.
- Network access control for outbound and inbound can be individually enabled or disabled.
- TCP/IP caches security information following network access control checks. Access is automatically rechecked on the next use of the cache whenever the RACF commands SETROPTS RACLIST, SETROPTS RACLIST REFRESH, or SETROPTS NORACLIST are issued for the SERVAUTH class. If you are using a security product other than RACF, the NetAccess zone table in the TCP/IP profile must be rebuilt to cause TCP/IP to recognize changes to the SERVAUTH class profiles for existing sockets.

  You can control the level of caching that is performed by using the CACHEALL, CACHEPERMIT, or CACHESAME parameter on the NETACCESS configuration statement in the TCP/IP profile. These parameters determine which network access control checks are resolved using the cache, without making a new call to SAF, and thus determine which checks can be audited by the security server product. For more information about these parameters and the NETACCESS statement, see z/OS Communications Server: IP Configuration Reference.
- The socket calls bind, connect, and those that send data will fail with errno EACCES if the specified IP address is not permitted to be used by the application user. Inbound UDP and RAW datagrams that are not permitted under the current network access control policy are normally filtered out before they get to socket calls that receive data. It is possible for a datagram to arrive under a network access control policy that would allow it to be read, but then be received after a policy change that does not allow it. If

the application (or common INET) has issued a select or poll on the socket, the receive call returns an EACCES errno to avoid blocking the application. Inbound TCP connections that are not permitted under the current network access control policy are also normally reset and discarded before they get to the socket backlog. In those cases where the only available new connections are not allowed and a select has been issued, accept processing returns a new socket and the next attempt to send or receive data returns an ECONNRESET error.

- NetAccess inbound filtering needs to predict whether a future receive or accept will succeed. When there are multiple processes or threads operating on the same socket, to achieve consistent results you must ensure that certain calls are done under the same identity, or identities with equivalent network access control policies. For UDP and RAW sockets, the select, receive, and send calls must be done under equivalent policies. For TCP listening sockets, the select and accept calls must be done under equivalent policies.

Figure 29 on page 142 provides an overview of network access control. z/OS user Bob is permitted access to Security Zone A but not Security Zone B. An outbound connect from Bob is permitted to Security Zone A, but not Security Zone B. Bob is permitted to accept connections from Security Zone A but not Security Zone B.

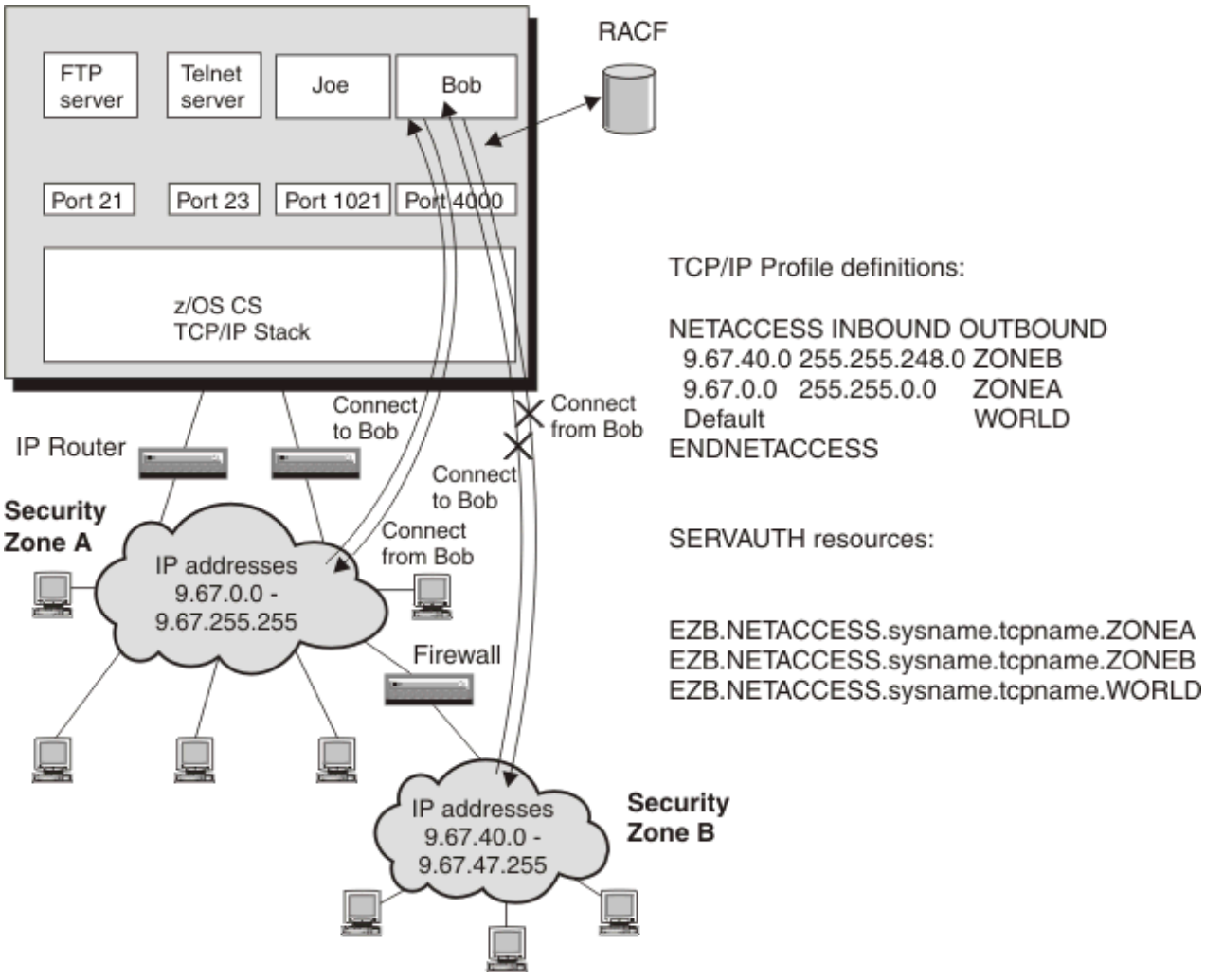


Figure 29. Network access control example

OSM access control

The intranode management network is intended for only authorized applications, such as those performing platform performance management functions. For more information about these applications, see *IBM z Systems® Ensemble Planning Guide*.

The intranode management network can be accessed only through OSM interfaces. To use IOCTL calls to retrieve information about an OSM interface or to send or receive data over OSM interfaces on this network, an application must have READ authorization to the EZB.OSM.*sysname*.*tcpname* resource. If you start one of these authorized applications on a z/OS image, authorize the application user ID to this SAF resource. In addition, authorize to this resource any user IDs that might issue diagnostic commands, such as Ping and Traceroute, over OSM interfaces to verify connectivity. Traffic over OSM interfaces is exempt from network access control.

For more information about the intranode management network, see Chapter 9, "TCP/IP in an ensemble," on page 493.

Socket option access control

Socket option access control gives system administrators the ability to assign permission for z/OS users to set selected socket options using a SAF-compliant security server. Access control is provided for the SOL_SOCKET level, SO_BROADCAST option, and the IPv6 advanced socket API options.

SO_BROADCAST socket option

IPv4 IP addresses are divided into a network portion, an optional subnetwork portion, and a host portion. Normally, UDP and RAW datagrams are delivered to the single peer system identified by the destination IP address. However, when the destination address is a subnetwork or network base address (host portion is all zeros), or a subnetwork or network broadcast address (host portion is all ones), the datagram is delivered to every peer system in that subnetwork or network. Socket semantics require that an application set the SO_BROADCAST option on before attempting to send a datagram to a base or broadcast address. This protects the application from accidentally sending a datagram to many systems. Network access control does not check to see if there are other security zones defined within the scope of a destination subnetwork or network address or whether the user is permitted to send a datagram to all of these security zones.

Control over setting this socket option allows the system administrator to restrict use of subnetwork or network addresses to those users or programs that require them. This support is enforced at the PFS layer and applies to all z/OS Communications Server socket APIs.

TCP/IP programs known to set the SO_BROADCAST socket option include:

- OMPROUTE
- sntpd, when invoked with the -b option

Additionally, any programs that use the clnt_broadcast() service in the SUN RPC libraries, or the send_pkt(sock, pkt, addr, broadcast) service in the NCS RPC library with the broadcast parameter set, require permission to the SO_BROADCAST socket option. The following TCP/IP programs use RPC services that require permission to broadcast:

- rpcinfo, when invoked with the -b option
- orpcinfo, when invoked with the -b option

The socket option to be protected is represented by the resource name EZB.SOCKOPT.*sysname*.*tcpname*.SO_BROADCAST. When this profile is defined, users of any program setting this option require READ permission. Access to the option is also allowed if the security server indicates there is no profile covering this resource. Conditional access lists, such as PERMIT WHEN(PROGRAM(...)), are supported for profiles covering socket option access control resources. There are no new TCP definitions required.

Guideline: Some security products do not distinguish between a resource profile that is not defined and a user that is not permitted to that resource profile. If your product does not make this distinction, you must define the socket option resource profile and permit users to it whenever the SERVAUTH class is active.

The following example shows the definitions:

```
RDEFINE SERVAUTH EZB.SOCKOPT.*.*.SO_BROADCAST UACC(NONE)
PERMIT EZB.SOCKOPT.*.*.SO_BROADCAST CLASS(SERVAUTH) ACCESS(READ) ID(OMPROUT)
```

```
PERMIT EZB.SOCKOPT.*.*.SO_BROADCAST CLASS(SERVAUTH) ACCESS(READ) ID(*)
WHEN(PROGRAM(ORPCINFO))
```

The program name listed in the conditional access list must be the name the program is invoked by. Most TCP/IP applications are invoked by an alias name rather than the module name. Table 11 on page 144 lists TCP/IP applications that send broadcast datagrams:

Table 11. TCP/IP application load module and alias names

Load module	**Alias**
EZAORRTE	OMPROUTE
EZASNTPD	SNTPD
EZARPCIN	RPCINFO
EZARORNP	ORPCINFO

Tip: In the UNIX System Services environment, both /bin/rpcinfo and /bin/orpcinfo are externally linked to ORPCINFO. Either command executes the EZARORNP program.

To use program names in conditional access lists, the program must be loaded into a controlled environment from a program controlled data set. TCP/IP applications are distributed in the SEZALOAD load library. To program control this data set, you must add it to the ** profile in the PROGRAM class as follows:

```
RALTER PROGRAM ** ADDMEMBER('TCPIP.SEZALOAD'//NOPADCHK)
```

For more information on program control, see z/OS Security Server RACF Security Administrator's Guide.

IPv6 advanced socket API options

You can control access for the IPv6 advanced socket API options that influence outbound packets.

For the IPV6_NEXTHOP, IPV6_TCLASS, IPV6_RTHDR, IPV6_HOPOPTS, IPV6_DSTOPTS, IPV6_RTHDRDSTOPTS, and IPV6_PKTINFO socket options, to set the socket option on setsockopt() or to use the corresponding ancillary data item on sendmsg(), an application must meet one of the following criteria:

- Be APF authorized.
- Have superuser authority.
- The corresponding SERVAUTH resource name in Table 12 on page 144 is defined, and the application has at least READ access to the resource.

Table 12. Socket option resource names

Socket option/Ancillary data item	**Resource name**
IPV6_NEXTHOP	EZB.SOCKOPT.*sysname.tcpname*.IPV6_NEXTHOP
IPV6_TCLASS	EZB.SOCKOPT.*sysname.tcpname*.IPV6_TCLASS
IPV6_RTHDR	EZB.SOCKOPT.*sysname.tcpname*.IPV6_RTHDR
IPV6_HOPOPTS	EZB.SOCKOPT.*sysname.tcpname*.IPV6_HOPOPTS
IPV6_DSTOPTS	EZB.SOCKOPT.*sysname.tcpname*.IPV6_DSTOPTS
IPV6_RTHDRDSTOPTS	EZB.SOCKOPT.*sysname.tcpname*.IPV6_RTHDRDSTOPTS
IPV6_PKTINFO	EZB.SOCKOPT.*sysname.tcpname*.IPV6_PKTINFO
IPV6_HOPLIMIT	EZB.SOCKOPT.*sysname.tcpname*.IPV6_HOPLIMIT

For the IPV6_HOPLIMIT socket option, to set a hop limit greater than the default using either the IPV6_UNICAST_HOPS or IPV6_MULTICAST_HOPS socket option on setsockopt() or the IPV6_HOPLIMIT ancillary data item on sendmsg(), an application must meet one of the following criteria:

- Be APF authorized.
- Have superuser authority.
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPLIMIT is defined, and the application has at least READ access to this resource.

TCP/IP applications that set IPv6 advanced socket API options

Table 13 on page 145 lists the TCP/IP applications that set IPv6 advanced socket API options. In a multilevel secure environment, APF and superuser authority are not sufficient; the security product resource name for the socket options appearing in Table 13 on page 145 must be defined for every stack. Usage must be explicitly permitted by user ID, or conditional access lists, such as PERMIT WHEN(PROGRAM(...)) must be defined for the load module and alias names listed in the table. For more information about applications in a multilevel secure environment, see "Required configuration in a multilevel secure environment" on page 195.

Table 13. TCP/IP applications that set IPv6 advanced socket API options

Load module	Alias	Socket options set
EZACDOPN	z/OS UNIX **ping**	IPV6_PKTINFO, IPV6_DONTFRAG
EZACDTPN	TSO PING	IPV6_PKTINFO, IPV6_DONTFRAG
EZACDTRT	z/OS UNIX **traceroute**	IPV6_HOPLIMIT, IPV6_PKTINFO
EZACDTTR	TSO TRACERTE	IPV6_HOPLIMIT, IPV6_PKTINFO
EZADNSVR	NAMED	IPV6_PKTINFO
EZAORRTE	OMPROUTE	IPV6_HOPLIMIT, IPV6_PKTINFO
EZASNTPD	SNTPD	IPV6_PKTINFO

For more details on these IPv6 advanced socket API options, see z/OS Communications Server: IPv6 Network and Application Design Guide.

For examples of the security product definitions, see the EZARACF sample in data set SEZAINST.

Netstat access control

You can control access to Netstat command output from the TSO or UNIX System Services shell environments using the System Authorization Facility (SAF). There are no new TCP definitions required. The Netstat command output is considered the resource to be protected, and you protect it by defining EZB.NETSTAT.*sysname*.*tcpname*.*netstat_option* resource profiles in the SERVAUTH class. A user ID has access to the Netstat output for a particular option if a security profile is defined in the SERVAUTH class for that option, and that user ID has READ access to the associated resource. Access is also given if there is no resource profile defined for a resource.

Guideline: Some security products do not distinguish between a resource profile that is not defined and a user that is not permitted to that resource profile. If your product does not make this distinction, you must define the Netstat resource profiles and permit users to them whenever the SERVAUTH class is active.

An installation can implement a security policy that indicates which users have authorization to selected Netstat options. The level of granularity for this security policy can be either by individual or all Netstat options.

Even though the Netstat RESCache/-q command returns system-wide resolver cache data, at least one TCP/IP stack must be active to ensure that the correct Netstat access controls are enforced for the

command. The same EZB.NETSTAT.*sysname.tcpname.netstat_option* naming convention applies to this Netstat command as it does for any other Netstat command.

Fast Response Cache Accelerator access control

Fast Response Cache Accelerator access control allows control of application access to Fast Response Cache Accelerator (FRCA) services. For more information on FRCA, see "Considerations for Fast Response Cache Accelerator" on page 56.

You can control application access to FRCA services by defining the EZB.FRCAACCESS.*sysname.tcpname* resource profile in the SERVAUTH class. Access to FRCA services is allowed if the web server user has READ access to this resource, or if the resource profile is not defined. There are no new TCP definitions required. This function is enabled if the SERVAUTH class is active and the FRCA resource profile is defined. If this resource profile is not defined, the check is not made.

Guideline: Some security products do not distinguish between a resource profile that is not defined and a user that is not permitted to that resource profile. If your product does not make this distinction, you must define the FRCA resource profile and permit users to it whenever the SERVAUTH class is active.

TCP/IP stack initialization access control

You can control whether an application can access a TCP/IP stack before the required policies have been installed, during the period of time after the stack is active but before Application Transparent Transport Layer Security (AT-TLS) has been initiated from policy. Access checking is performed only if you have configured AT-TLS in the initial PROFILE.TCPIP configuration data set. If AT-TLS is configured, you can control whether an application can access the TCP/IP stack before the required policies have been installed by defining the EZB.INITSTACK.*sysname.tcpname* resource profile in the SERVAUTH class. During the period of time after the stack is active and before AT-TLS has been initiated from policy, all socket requests are blocked if this resource profile is not defined. If this resource profile is defined, access to the stack is permitted only to user IDs that have READ access to the resource. Checking ceases the first time that the Policy Agent processing of the AT-TLS policy is completed, or when a profile change deactivates AT-TLS.

Guideline: Some security products do not distinguish between a resource profile that is not defined and a user that is not permitted to that resource profile. However, if AT-TLS is enabled, a profile that is not defined has the same result as a user that is not permitted, regardless of the security product you are using. If AT-TLS is enabled, you must activate the SERVAUTH class, and define the INITSTACK resource profile and permit users to it.

TCP/IP packet trace service access control

The TCP/IP packet trace service provides an interface for network management applications to obtain packet trace data. The information provided through the service is considered the resource to be protected. Access to this information can be controlled through an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname.tcpname*.SYSTCPDA.

Access to the information is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, it should be enabled on the stack using the NETMONITOR PKTTRCService statement in PROFILE.TCPIP. For details, see z/OS Communications Server: IP Configuration Reference.

If the resource profile is not defined, the service allows access to the packet trace information only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

TCP connection information service access control

The TCP connection information service allows network management applications to obtain information about TCP connection activity. Access to this information can be controlled by an

external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname.tcpname*.SYSTCPCN.

Access to the TCP connection information is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, it should be enabled on the stack using the NETMONITOR TCPCONNService statement in PROFILE.TCPIP. For details, see z/OS Communications Server: IP Configuration Reference.

If the resource profile is not defined, the service allows access to the TCP connection information only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

zERT information service access control

The z/OS encryption readiness technology (zERT) information service allows network management applications to obtain information about the cryptographic protection state of TCP and Enterprise Extender connections. Access to this information can be controlled by an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname.tcpname*.SYSTCPER.

Access to the zERT information is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, the ZERTService , ZERTServiceByPolicy, or both parameters must be specified on the NETMONITOR statement in the TCP/IP profile data set. For details, see z/OS Communications Server: IP Configuration Reference.

If the resource profile is not defined, the service allows access to the zERT information only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

Real-time SMF information service access control

The SMF information service allows network management applications to obtain selected TCP/IP SMF records, such as SMF records supported by FTP and Telnet, in a real-time fashion. Access to this information can be controlled through an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname.tcpname*.SYSTCPSM.

Access to these SMF records is allowed if the user ID associated with the network management application is permitted (read access) to this resource profile. In addition, to use this service, it should be enabled on the stack using the NETMONITOR SMFService statement in PROFILE.TCPIP. For details, see z/OS Communications Server: IP Configuration Reference.

If the resource profile is not defined, the service allows access to the SMF data only to superusers, or those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

TCP/IP OSAENTA trace service access control

The TCP/IP OSAENTA trace service is an interface that enables network management applications to obtain packet trace data. You should protect the information that is provided by this service. You can control access to this information through an external security manager product, such as RACF, by defining the SERVAUTH profile name EZB.NETMGMT.*sysname.tcpname*.SYSTCPOT.

Access to this information is allowed if the user ID associated with the network management application is permitted to (has read access to) to this resource profile. To use this service, ensure that it has been enabled on the stack using the NTATRCService parameter on the NETMONITOR statement in PROFILE.TCPIP. For details about the NETMONITOR statement, see z/OS Communications Server: IP Configuration Reference.

If you do not define this resource profile, the service allows access to the packet trace information only to a superuser, or to those permitted to become a superuser (those with read access to BPX.SUPERUSER).

IPSec network management interface access control

The IPSec network management interface (NMI) enables network management applications to obtain detailed information for and exercise control over IP filtering and IPSec security associations. Access to this interface can be controlled through an external security manager product, such as RACF, by defining the SERVAUTH profile names EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.DISPLAY and EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.CONTROL for display requests and control requests respectively.

Applications can access this interface if the user ID associated with the network management application is permitted (has read access) to the appropriate resource profile.

If the resource profile is not defined, the service allows access to the IPSec NMI only to superusers, or to those permitted to become superusers (that is, those with read access to BPX.SUPERUSER).

For more information about the IPSec NMI, see z/OS Communications Server: IP Programmer's Guide and Reference.

CIM provider access control

CIM provider access control permits the Communications Server CIM providers to gather CIM data, when the user ID associated with the client of the z/OS CIM server is not defined as a superuser. For more information on the CIM providers, see "Considerations for Common Information Model providers" on page 109.

Access can be controlled by an external security manager product, such as RACF, by defining the resource profile name EZB.CIMPROV.*sysname*.*tcpname* in the SERVAUTH class. For examples of the security product commands needed to create this resource profile name and grant users access to it, see member EZARACF in sample data set SEZAINST.

Access is granted if the user ID associated with the client of the z/OS CIM server is permitted (has read access) to this resource profile.

Guideline: Some security products do not distinguish between a resource profile that is not defined and a user that is not permitted to that resource profile. If your product does not make this distinction, you must define the CIM provider resource profile and permit the client user ID to it whenever the SERVAUTH class is active, if you want the Communications Server CIM providers to be able to gather CIM data.

Real-time application-controlled TCP/IP trace NMI access control

The real-time application-controlled TCP/IP trace NMI provides real-time trace and event information to network management applications. For applications to use some of the functions of this NMI, you must define the following resource profiles and give the application user ID read access to the associated resources. If profiles that protect these resources are not defined, no applications are authorized to use the associated resources.

- If you want to permit network management applications to open a trace instance, define a profile for the following resource:

  ```
  EZB.TRCCTL.sysname.tcpname.OPEN
  ```

- If you want to permit network management applications to set packet trace filters and options, define a profile for the following resource:

  ```
  EZB.TRCCTL.sysname.tcpname.PKTTRACE
  ```

- If you want to permit network management applications to obtain packet trace IPSec cleartext data, define a profile for the following resource:

  ```
  EZB.TRCSEC.sysname.tcpname.IPSEC
  ```

- If you want to permit network management applications to set data trace filters and options, define a profile for the following resource:

```
EZB.TRCCTL.sysname.tcpname.DATTRACE
```

- If you want to permit network management applications to obtain data trace AT-TLS cleartext data, define a profile for the following resource:

```
EZB.TRCSEC.sysname.tcpname.ATTLS
```

For more information about the real-time application-controlled TCP/IP trace NMI, see z/OS Communications Server: IP Programmer's Guide and Reference.

Syslogd isolation

Syslogd isolation provides a capability for the installation to control which user IDs and job names can write syslogd records to specified syslogd facilities. This function enables the installation to separate system and application syslogd records, and to separate syslogd records from different applications. This function prevents an application level process from flooding a syslogd facility intended for system use, possibly causing system syslogd records to be lost. This function is enabled when user ID and/or job name are specified as additional criteria along with existing facility and priority criteria to select a syslogd repository.

In addition, the user ID and job name associated with the syslogd record writer can optionally be stored in a syslogd record based on a syslogd command-line parameter. This capability is useful when syslogd records for multiple jobs or users are recording in the same syslogd facility. This function enables positive identification of the creator of the syslogd records and ensures that the syslogd record, if spoofed, can be identified.

Syslogd isolation also provides a capability to disable reception of syslogd messages from other hosts in the network. This capability is provided by a syslogd command-line parameter. This parameter disables reception of syslogd messages from all hosts. If an installation wants to allow certain hosts in the network access to syslogd, IP Filtering can be used instead to specify which hosts are permitted to access the syslogd UDP port.

IP filtering

The IP security function can configure the Communications Server to perform packet filtering at the IP layer for IPv4 and IPv6.

IP filters are rules defined to either discard or permit packets. IP filtering matches a filter rule to data traffic based on any combination of IP source or destination address (or masked address), protocol, source or destination port, direction of flow, or time. IP filtering can control traffic being routed, or control access at the host that has the communication endpoint. Even when an external firewall is providing filtering protection for the host, Communications Server IP filtering can provide a secondary line of defense.

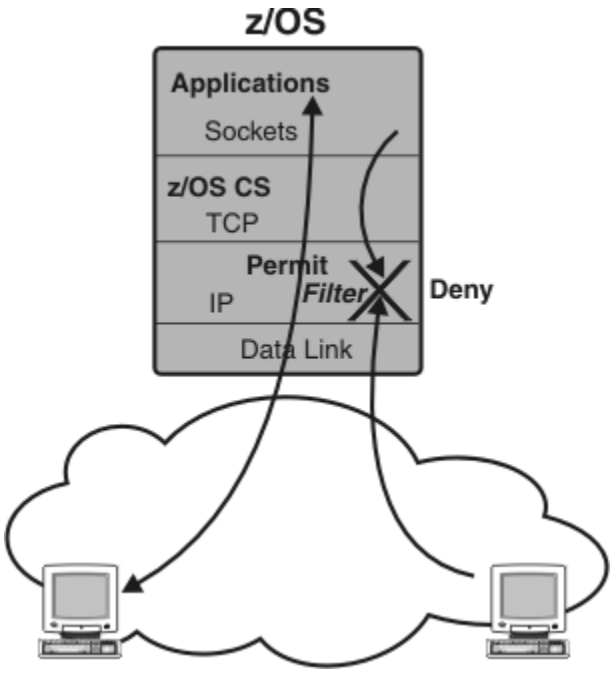


Figure 30. IP filtering at the z/OS communication endpoint

For more information about IP filtering, see Chapter 17, “IP security,” on page 883.

Security considerations for the VARY command

You can restrict access to the VARY TCPIP command by defining RACF profiles under the OPERCMDS class and specifying the list of users that are authorized to issue the VARY TCPIP command. You can decide on the level of control that is appropriate for your installation. For example, you might want to allow a user to be able to start or stop a TCP/IP device using the VARY TCPIP command, but you do not want the user to be able to modify the TCP/IP configuration. For further information on restricting access to the VARY TCPIP command, see z/OS Communications Server: IP System Administrator's Commands.

Multilevel security

Multilevel security is an enhanced security environment that can be configured on a z/OS system. In this environment, the security server and trusted resource managers enforce mandatory access control policies in addition to the usual discretionary access control policies. To participate in a multilevel-secure environment, the user IDs associated with tasks trying to access z/OS CS resources and those resource profiles in the SERVAUTH class need to have security labels defined. For more information on the multilevel-secure environment and configuring z/OS CS in that environment, see Chapter 4, “Preparing for IP networking in a multilevel secure environment,” on page 191.

Configuration data set security

The z/OS Communications Server configuration data sets can be encrypted by using z/OS data set encryption. You can specify key labels to identify encryption keys that are in the Cryptographic Key Data Set (CKDS). When an encrypted data set is created, the key label is stored as an attribute of the data set in the catalog. Authorization to view the contents of a data set is based on access to the key label that is associated with the data set. The encryption key that is associated with that key label is used by DFSMS to encrypt and decrypt the data as it is written or read. See *IBM Redbook: Getting Started with z/OS Data Set Encryption* for more information.

Note: You should consider very carefully when enabling data set encryption for resolver data sets accessed by applications issuing resolver calls and FTP client data sets accessed by z/OS FTP clients. In many cases, most users may require access to these data sets due to the wide use of the functions. In order to allow any applications that issue resolver calls and z/OS FTP clients to access these configuration

data sets with data set encryption enabled, it is necessary to give UACC(READ) to the key label associated with the resolver and FTP client data sets. See details in the following table.

Table 14. Configuration data sets

Configuration data set names	Description
ETC.IPNODES	One of the local host data sets used for IPv6 name query, or IPv4 and IPv6 name query when COMMONSEARCH is specified in the resolver setup file. See "Resolver configuration files" on page 771 for more information.
ETC.PROTO	Used to map types of protocol to integer values to determine the availability of the specified protocol. See "Resolver configuration files" on page 771 for more information.
ETC.SERVICES	Establishes port numbers for servers using TCP and UDP. See "Resolver configuration files" on page 771 for more information.
HOSTS.LOCAL	Input data set to MAKESITE for generation of HOSTS.ADDRINFO and HOSTS.SITEINFO. See "Resolver configuration files" on page 771 for more information.
HOSTS.SITEINFO and HOSTS.ADDRINFO	Local host data sets used for IPv4 name query when NOCOMMONSEARCH is specified in the resolver setup file. See "Resolver configuration files" on page 771 for more information.
FTP.DATA	Overrides default FTP client and server parameters for the FTP server. For more information, see File Transfer Protocol.
SOCKSCONFIGFILE	The FTP client uses this SOCKS server configuration data set to determine which FTP servers require SOCKS protocols. See File Transfer Protocol for more information.
TCPIP.DATA	Provides parameters for TCP/IP client programs. See Resolver setup and TCPIP.DATA configuration statements for more information.
TCPXLBIN	The translate tables (EBCDIC-to-ASCII and ASCII-to-EBCDIC) are referenced to determine the translate data sets to be used.

Network security principles

This topic describes network security principles that you can use to protect data in your network.

Cryptography: The foundation of good security

The foundation of good security methods begins with cryptography. Cryptography keeps your data and communications secure using techniques such as encryption, authentication, and data integrity. Encryption services protect sensitive data from being read by other than the intended receiver. Cryptographic authentication and data integrity services allow communicating hosts to detect if data is altered in transit. Public key cryptography can identify and authenticate hosts or users. Public key cryptography can also be used in the secure creation of symmetric session keys for both security endpoints. After a secure session is created, successful data authentication and decryption occur only if both hosts have the correct session keys.

Cryptographic standards and FIPS 140

The National Institute of Standards and Technologies (NIST) publishes Federal Information Processing Standards publication 140 (FIPS 140). This publication specifies security requirements for cryptographic modules for both hardware and software components of computer systems. FIPS 140 places some restrictions on the use of cryptographic algorithms and modules. Some examples of the restrictions are:

- Cryptographic algorithms and keys must be contained within a cryptographic module and accessed through a well defined cryptographic boundary.
- Use of weaker cryptographic algorithms (for example, DES and MD5) is not allowed.
- Use of weaker asymmetric key lengths (for example, RSA digital signature operations using key lengths less than 1024 bits) is not allowed.
- Use of Diffie-Hellman groups with weaker key lengths (key lengths less than 2048 bits) is not allowed. This restriction applies to groups 1, 2, and 5.

See the National Institute of Standards and Technology (NIST) website at http://csrc.nist.gov/publications/PubsFIPS.html for the most recent FIPS 140 publication, and other related publications.

On z/OS systems, Integrated Cryptographic Services Facility (ICSF) and System SSL provide cryptographic services. z/OS Communications Server uses ICSF and System SSL in addition to its own cryptographic algorithms in some of its networking security functions, such as AT-TLS and IP security. You can configure ICSF, System SSL, and the z/OS Communications Server networking security functions in FIPS 140 mode, in which case they enforce FIPS 140 restrictions. Enabling FIPS 140 mode might require additional setup and configuration, and it might result in a reduction in performance.

See the following references for information about configuring z/OS functions in FIPS 140 mode:

- "FIPS 140 mode and IP security" on page 889
- Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121
- z/OS Cryptographic Services System SSL Programming
- z/OS Cryptographic Services ICSF Overview
- z/OS Cryptographic Services ICSF Administrator's Guide
- Operating in compliance with FIPS 140-2 in z/OS Cryptographic Services ICSF Writing PKCS #11 Applications

End to end security

Cryptographic security solutions can be applied to a portion of the data path or end to end, whichever is appropriate for your security policy. Generally, the greatest degree of security is provided when cryptographic methods are used end to end. However, if only portions of the data path are considered untrusted by an enterprise (such as the Internet) it may be adequate to protect only the untrusted portion with cryptography. z/OS offers security protocols that can be configured to protect portions of the data path or the entire data path.

Workload-based security deployment

In making a security protocol selection, an important consideration is the application workload to be protected. In order to illustrate this concept, it is helpful to understand where various protocols are implemented from a protocol layering perspective.

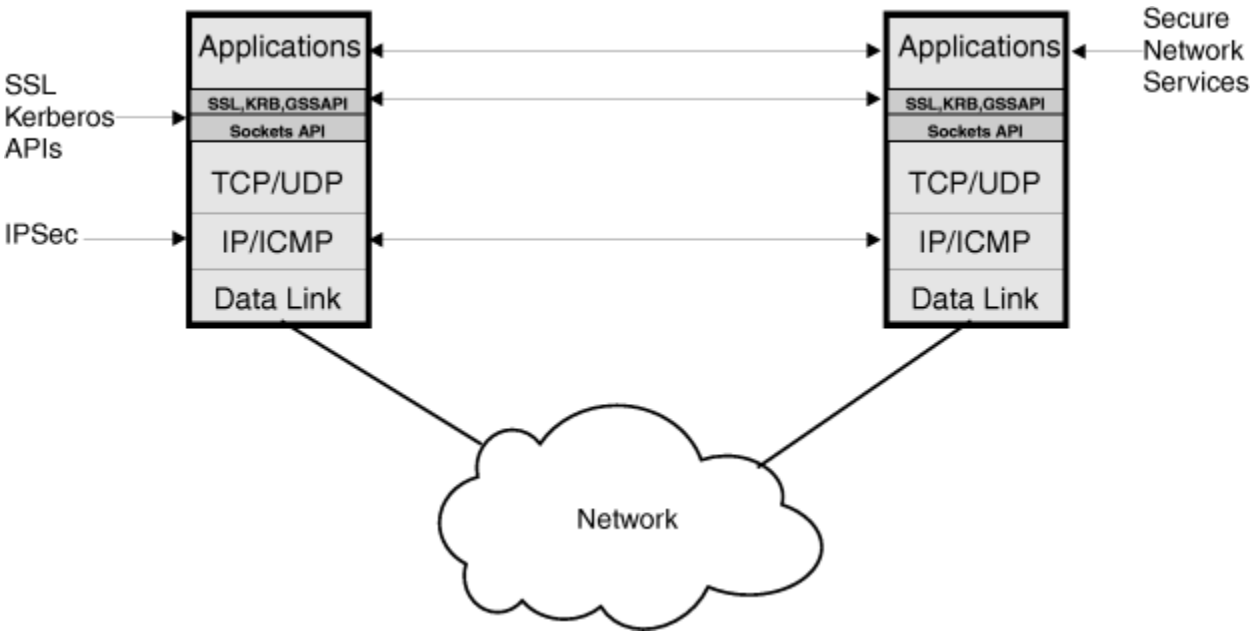


Figure 31. Security protocols from a protocol layering perspective

Existing workload

The network layer is the lowest layer in the protocol stack where end to end security over multiple hops can be applied. Network layer security protocols provide blanket protection for upper-layer application data without requiring modification to the application. IPSec is implemented at the network layer and provides authentication, integrity, and data privacy between any two IP entities. IPSec can protect a segment of the data path (e.g., between two routers), or it can secure the data path end to end. Because IPSec is applied at the IP layer, it is a connectionless security protocol and is applied on a per packet basis.

Transport Layer Security (TLS), which evolved from secure Sockets Layer (SSL), is another popular security protocol implemented above the transport layer at the application interface layer. TCP applications can be modified to use TLS. Many existing socket applications might be able to use Application Transparent Transport Layer Security (AT-TLS) services provided within the TCP/IP TCP layer. For details, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

TLS requires a reliable transport layer and is therefore not used for UDP applications. TLS provides authentication, integrity, and data privacy. TLS is a connection-oriented security protocol and protects all data on a connection or session.

The Communications Server has a TLS-enabled TN3270E Telnet server, which provides secure access to existing SNA applications being accessed over an IP network. Serving as a protocol gateway between the IP network and the SNA network, the TLS-enabled TN3270E Telnet server protects the data path in the IP network from the Telnet client all the way to the z/OS TN3270E Telnet server. If the TN3270E Telnet server is on a different host from the target SNA application, IPsec can be used to secure Enterprise Extender links in the SNA portion of the data path. Enterprise Extender application data can be protected without modification to the SNA applications.

New workload

For new applications, security can be built-in. One method of building security into the application on z/OS is to use z/OS System SSL and Kerberos. Another method is to use the Communication Server's policy-driven Application Transparent Transport Layer Security (AT-TLS), which requires no change to the application unless the application must control certain portions of the AT-TLS support.

Tip: Using AT-TLS is highly recommended over the direct use of System SSL.

Newer versions of network services, such as SNMPv3 that are supported by the Communications Server, have security built into the application protocol using standards-based specifications for secure interoperability.

Network security protocols

This topic describes network security protocols that you can use to protect data in your network.

IPSec and VPNs

IPSec is defined by the IPSec Working Group of the IETF. It provides authentication, integrity, and data privacy between any two IP entities. Management of cryptographic keys and security associations can be done manually or dynamically using an IETF-defined key management protocol called Internet Key Exchange (IKE).

There are two versions of the IKE protocol:

- IKE version 1.0 (IKEv1) is defined by RFC 2409, *The Internet Key Exchange (IKE)*, and related RFCs. This is the version that has been supported by z/OS Communications Server for a number of years.
- IKE version 2.0 (IKEv2) is defined by RFC 5996, *Internet Key Exchange Protocol: IKEv2*, and related RFCs. Support for IKEv2 is introduced with z/OS V1R12.

With IPSec, you can create virtual private networks (VPN). A VPN enables an enterprise to extend its private network across a public network, such as the Internet, through a secure tunnel called a security association. IPSec VPNs enable the secure transfer of data over the public Internet for same-business and business-to-business communications, and protect sensitive data within the enterprise's internal network.

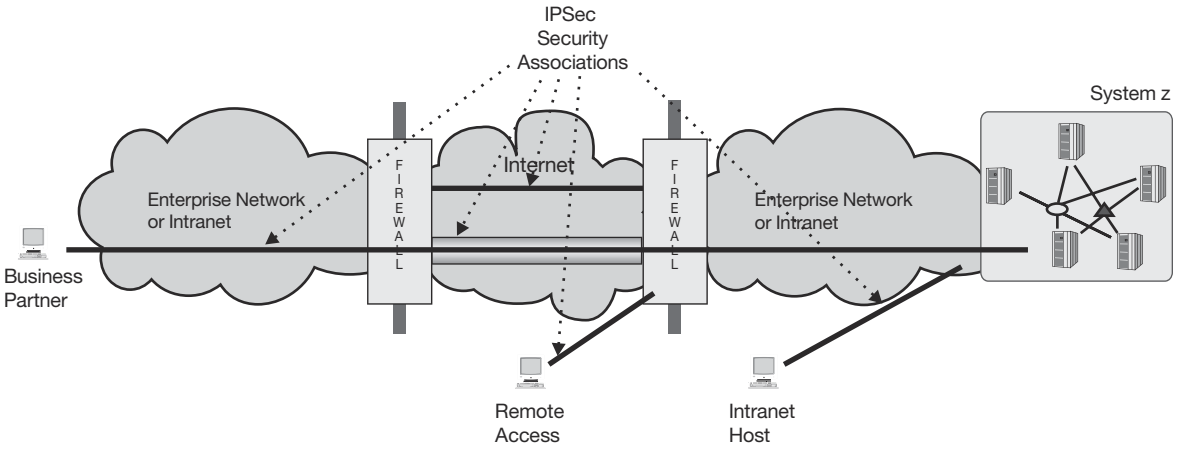


Figure 32. e-business scenarios with virtual private networks

z/OS provides support for IKE and IPSec VPNs, including the following options:

- AH and ESP protocols
- Triple DES
- AES with several choices of mode or key length
- IPSec transport and tunnel mode encapsulation
- IKEv1 and IKEv2 negotiations with support for both aggressive and main mode in IKEv1
- Pre-shared key and digital signature methods of authentication
- NAT traversal (IPv4 only)

For more information about configuring IPSec and VPNs, see Chapter 17, “IP security,” on page 883.

For more information on using IPSec with Dynamic VIPAs, see “Sysplex-Wide Security Associations” on page 396.

Hardware features for encryption, decryption and hashing

IBM CP Assist for Cryptographic Functions (CPACF), where available, or a cryptographic coprocessor provide support for IPSec encryption, decryption, and hashing functions. Where such hardware features are not available, IPSec's encryption, decryption, and hashing functions are performed in software.

Additional IPSec assist using System z Integrated Information Processor (zIIP IP security)

On a System z9 or later server, an additional assist for IPSec protocol traffic is available with the System z Integrated Information Processor (zIIP). To enable zIIP IP security in Communications Server, specify ZIIP IPSECURITY on the GLOBALCONFIG statement. With zIIP IP security enabled, traffic using the AH and ESP protocols can be processed on available zIIPs. When enabled on a z9 or later z/OS image that includes zIIPs, the zIIP IP security function can reduce the IPSec processing load on general purpose central processors, beyond what is achievable using just CPACF or the Cryptographic Coprocessor.

When zIIP IP security is enabled, you might have to modify some Workload Manager (WLM) definitions. The IPSec traffic that can be processed on available zIIP processors is assigned to an independent WLM enclave. The WLM independent enclave encapsulates the IPSec workload as execution units that are separately classified and managed in a WLM service class. See the following considerations:

- All WLM independent enclaves that are not classified are assigned WLM service class SYSOTHER (with a goal of discretionary). In many cases, enclave workload using this service class can result in performance degradation if the workload is left unclassified in the service definition. Also, service that is accumulating in the SYSOTHER service class is an indication that you have unclassified workload in your system.
- Examine the IIPHONORPRIORITY parameter located in the IEAOPT*xx* member of SYS1.PARMLIB. When this parameter is specified as NO, general purpose central processors do not process zIIP-eligible workload when zIIPs are online. Omitting the IIPHONORPRIORITY parameter or specifying IIPHONORPRIORTY=YES allows zIIPs to request help from general-purpose central processors when the zIIPs can not complete all zIIP-eligible workload within a reasonable period of time (see ZIIPAWT in z/OS MVS Initialization and Tuning Reference).
- If there are any other workloads that are eligible to run on your zIIPs, analyze your current WLM workload goals and make any necessary adjustments. For example, if you are operating a workload that is more latency-sensitive than the typically longer running IPSec workloads, such as Db2® Distributed Relational Database Architecture™ (DRDA) workload, consider classifying the IPSec workload to make it less preferable than the workload that is more latency-sensitive.

  Guideline: Make these two performance goal settings for the IPSec independent enclave:

  – Set the WLM service class associated with the IPSec independent enclave to a lower execution velocity goal than that which is being assigned to the more latency-sensitive workload (such as Db2 DRDA). The lower execution velocity goal is chosen because the IPSec independent enclave has no associated transactions.

  – Set the WLM service class associated with the IPSec independent enclave to a greater importance level value (importance is defined in five levels, 1 to 5, with 1 indicating highest importance) than that which is being assigned to the more latency-sensitive workload (such as Db2 DRDA). Each WLM service class is associated with an importance level that specifies how important it is to your business that this workload is meeting its goal. The importance level defines how work is treated by the system. Achieving the velocity goal for IPSec traffic that can be processed on available zIIP processors is less important than the more latency-sensitive workload.

Because an independent enclave enables WLM to manage the priority of all workload in the enclave, you should classify the workload for IPSec traffic. To classify the independent enclave used for IPSec workload, make the following WLM service definitions using the WLM ISPF panels:

1. Create a workload for the IPSec traffic that will be operating on the independent enclave.

   From the primary WLM ISPF panel, select option 2 `Workloads`.

2. Create a service class that contains an appropriate performance goal for the IPSec independent enclave.

   On the primary WLM ISPF panel, select option 4 `Service Classes`. From this panel, define your new service class and associate it with the workload you previously defined. When you define the BASE GOAL information for your single defined period, choose the goal type `Execution velocity`. After this is selected, define a velocity and importance for the service class that you are defining. Set a value that takes into account other traffic that might be competing for zIIP or general central processor resources. (General central processors become a factor when you have set the IIPHONORPRIORITY parameter to the value YES in the IEAOPT*xx* member of SYS1.PARMLIB.)

3. Create a WLM subsystem type for TCP/IP.

   You must specify the subsystem type name as TCP; define it by using the WLM ISPF application. On the primary WLM ISPF panel, select option 6 `Classification Rules`; the `Subsystem Type Selection List for Rules` panel is displayed. Move your cursor to the field `Subsystem-Type` and press the Enter key. When you are prompted for the type of operation that you want to perform, select option 1 `Create`, because you want to create a new subsystem type. On the `Create Rules for the Subsystem Type` panel, specify the subsystem type TCP and a description for this new subsystem type.

4. Create a classification rule for the subsystem type TCP on the `Create Rules for the Subsystem Type` panel of the WLM ISPF application.

   Define a classification rule for the subsystem type. This rule determines what workload is associated with a service class for this subsystem type. You can use the following workload qualifiers for the new independent enclave for IPSec workload:

   - Subsystem Instance (SI) will be set to the job name of the TCP/IP stack
   - Transaction Name will be set to a value of TCPENC01

To verify that the new independent enclave is being used with an appropriate WLM service class, use the System Display and Search Facility (SDSF) ENC command or view the RMF Monitor III ENCLAVE report (or use any other method to interactively view RMF data).

For a more detailed description of defining Workload Manager (WLM) service definitions (workloads, service classifications, classification rules, subsystem type, and so on) and WLM in general, see *System Programmer's Guide to: Workload Manager* (IBM Redbooks) and z/OS MVS Planning: Workload Management. For information about configuring the IIPHONORPRIORITY parameter in the IEAOPT*xx* member of SYS1.PARMLIB, see z/OS MVS Initialization and Tuning Reference. For more information about viewing enclaves using SDSF, see z/OS SDSF Operation and Customization. For additional information about the RMF workload activity report, see z/OS RMF Report Analysis.

On system models with no zIIPs (z990, or a z9 or later with no zIIPs configured), you can enable zIIP IP security so that you can project the percentage of existing IPSec workload (running on central processors) that would be eligible to run on zIIPs, if zIIPs were available on the z/OS image. To perform projection analysis, specify ZIIP IPSECURITY on the GLOBALCONFIG statement, and specify PROJECTCPU=YES in the IEAOPT*xx* member of SYS1.PARMLIB. Run your IPSec workload, and SMF provides accounting information regarding workload that is eligible to run on zIIPs. For information about configuring the PROJECTCPU parameter in the IEAOPT*xx* member of SYS1.PARMLIB, see z/OS MVS Initialization and Tuning Reference. For information about accounting for zIIP eligibility in SMF record types 30 and 7*x*, see z/OS MVS System Management Facilities (SMF). For information about zIIP-related reporting updates, see z/OS RMF Report Analysis.

Guidelines:

- Because cryptographic hardware performance differs significantly between z9 or later processors and processors that preceded the z990, you should not use zIIP IP security for projection purposes on processors preceding the z990.
- TCP/IP consumes slightly more central processing resources when no zIIPs are online and you have coded GLOBALCONFIG ZIIP IPSECURITY. Remove GLOBALCONFIG ZIIP IPSECURITY from your TCP/IP profile after you have completed your zIIP performance projection runs.

TLS and SSL

The Transport Layer Security (TLS) protocol provides data encryption, data origin authentication, and message integrity. It also provides server and client authentication using X.509 certificates. TLS begins with a handshake during which the server is authenticated to the client using X.509 certificates. Optionally, the client can also be authenticated to the server using X.509 certificates. During the handshake, the TLS protocol version and security session parameters, including suites of cryptographic algorithms called cipher suites, are negotiated and session keys are created. After the handshake, the data is protected during transmission with data origin authentication, integrity and encryption using the session keys.

The cryptographic algorithms that are used for the TLS session are based on the cipher suite that the server and client negotiate. During the TLS handshake, the client and server exchange a list of cipher suites. The suite that is selected is based on the best match between the client list and the server list. You can limit the selectable algorithms by configuring a subset of allowable algorithms at the server. TLS supports cipher suites that include AES-based encryption and a variety of other encryption algorithms. Cryptographic hardware features like CPACF and Crypto Express adapters, if available, are used to accelerate certain cryptographic operations.

TLS typically requires a server X.509 certificate and associated private key, which are stored in a keystore such as a SAF key ring, a gskkyman key database, or a Java™ keystore. The certificate is used as part of the TLS handshake server authentication process. The client validates the server certificate. TLS optionally uses a client X.509 certificate that is used as part of the TLS handshake client authentication process. In order to use client authentication, the client must have a client X.509 certificate and associated private key. Successful client authentication requires that the Certificate Authority (CA) that signed the client certificate be considered trusted by the server. To be considered trusted, the certificate of the CA must be in the key store of the server.

See “Transport Layer Security” on page 598 for detailed information on obtaining certificates.

TLS is based on the Secure Sockets Layer (SSL) protocol and is defined by the Internet Engineering Task Force (IETF) in RFCs 2246 (TLSv1.0), 4346 (TLSv1.1), and 5246 TLSv1.2) and 8446 (TLSv1.3). SSL was originally defined as a proprietary protocol, not by the IETF. Since TLS evolved from SSL, the two terms are used interchangeably throughout this book. When a specific TLS or SSL protocol version is intended, it will be specifically noted.

On z/OS there are two different TLS implementations. System SSL, a component of the Cryptographic Services element, provides a full set of TLS APIs for C and C++ programs. For Java programs, z/OS provides a full function Java Secure Sockets Extension (JSSE) provider. z/OS Communications Server's AT-TLS support uses System SSL for its TLS protocol processing.

TN3270E Telnet server security

The Communications Server provides z/OS TN3270E Telnet server (Telnet), that is enabled for AT-TLS; the data path in the IP network to Telnet is protected using the TLS protocol. IBM Host On Demand and Personal Communications provide a Telnet client that is enabled for TLS.

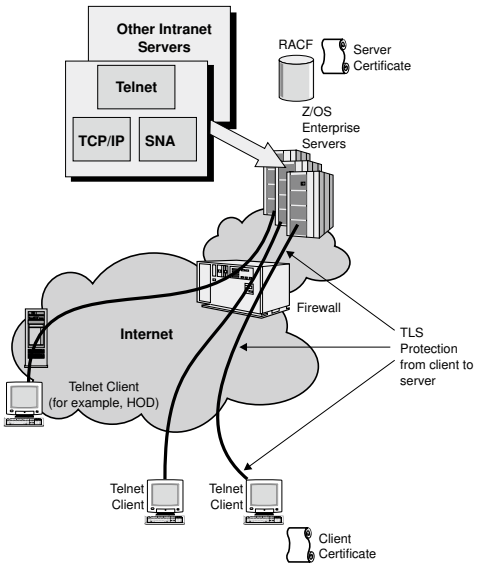


Figure 33. TN3270E Telnet server security overview

The Communications Server Telnet TLS support provides several extensions for RACF-based access control to Telnet. These extensions prevent a client from seeing the USSMSG (log on screen) unless the client is authorized. To use this support, define the client certificate to RACF using RACF digital certificate services. The first level of authorization checking verifies that the RACF user ID represented by the client certificate is defined to RACF. The next level of authorization requires that this RACF user ID be permitted to access the Telnet port. The Telnet port is represented as a RACF resource using the SERVAUTH class.

Multiple port support

You can use Telnet multiple port support to enable a combination of secure and non-secure traffic. To use multiple port support, you define separate ports; one port is dedicated to non-secure traffic and another port is dedicated to secure traffic. As an example, assume a scenario where Intranet client connections are not required to be secure, but Internet client connections are so required. In this case, Intranet clients connect to the BASIC port (port 23 in Figure 34 on page 158). Since all clients connecting from the Internet require secure connections, these clients use the TTLSPORT (port 1023 in Figure 34 on page 158). Packet filtering is used at the firewall that separates the intranet and the Internet to control access to the Telnet ports. To prevent Internet access to the BASIC port, port 23 is blocked at the firewall. The TTLSPORT, port 1023, is permitted at the firewall. In this scenario, the best security is achieved when TLS client authentication with the Telnet RACF extensions is used. This support ensures that the client has the authority to attempt to log on to SNA applications through Telnet. Regardless of the method of authentication used, the SNA application should identify and authenticate the user using RACF before any application access is granted. If you are using TLS encryption services, the user ID and password are encrypted.

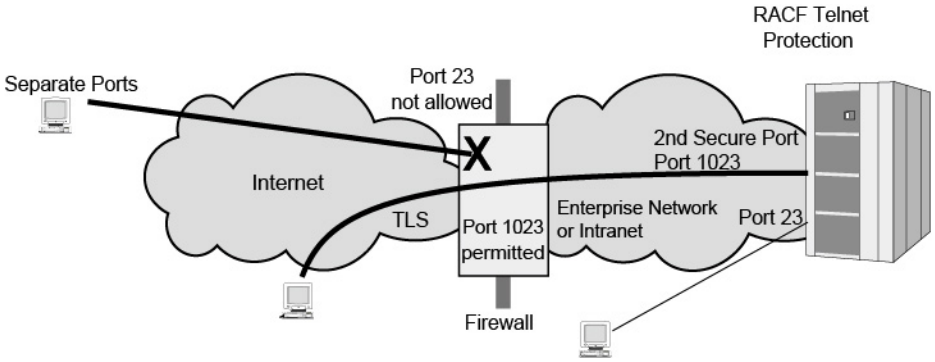


Figure 34. Using multiple Telnet ports to separate secure and non-secure traffic

Figure 35 on page 159 shows how you can combine IPSec and Telnet security to provide more secure remote access from the Internet to SNA applications than is depicted in Figure 34 on page 158. In this scenario, IPSec AH protocol is used for authentication between the user's PC and the firewall. The firewall is open for port 1023 for traffic that is authenticated with only IPSec. The firewall discards traffic for port 1023 that cannot be authenticated by IPSec. The additional security provided by IPSec protects the z/OS server from unauthorized access attempts and denial-of-service attacks by hosts outside the VPN.

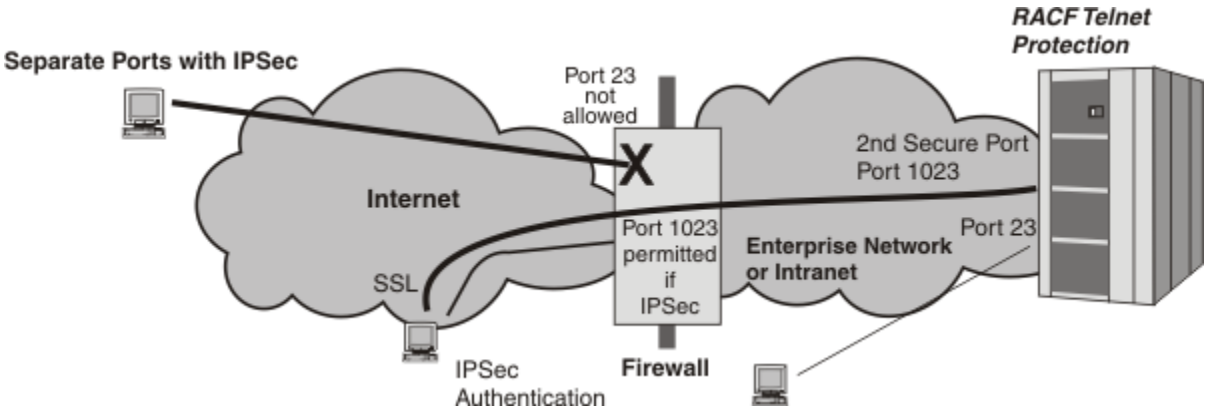


Figure 35. Combining Telnet security with IPSec client-to-firewall authentication

Secure and non-secure connections using a single Telnet port

A single port can be used to support a mix of secure and non-secure traffic. The port has the designation TTLSPORT. To support the configuration of various security policies for a single port, the TTLSPORT designation indicates that the port can use TLS, but the port does not have to use TLS.

Telnet supports both negotiated and non-negotiated TLS. Negotiated TLS is an IETF-defined extension to the TN3270 protocol. With negotiated TLS, the decision to use TLS for a connection is based on the outcome of a negotiation between the Telnet client and server using TN3270 protocols. This negotiation is performed after the Telnet connection is established, and if TLS is negotiated, the TLS handshake is performed. With non-negotiated TLS, a TLS handshake is required immediately after the connection is established. A single port can concurrently use both negotiated and non-negotiated TLS connections.

Figure 36 on page 160 shows a single Telnet port that allows a mix of secure and non-secure traffic. Intranet clients are not required to be secure. All clients connecting from the Internet are required to use TLS. Both intranet and Internet clients connect to the port designated as TTLSPORT (port 1023 in this example). In this scenario, IPSec AH protocol is used for authentication between the user's PC and the firewall. The firewall is open for port 23 for traffic that is authenticated with only IPSec. The firewall discards traffic for port 23 that IPSec cannot authenticate. In this scenario, packet filtering without IPSec cannot be used at the firewall that separates the intranet and the Internet to control access on the basis of port, because only one port is used. Without IPSec AH, all access control checks are deferred to Telnet. The additional security provided by IPSec at the firewall protects the z/OS server from unauthorized access attempts and denial-of-service attacks by hosts outside the VPN.

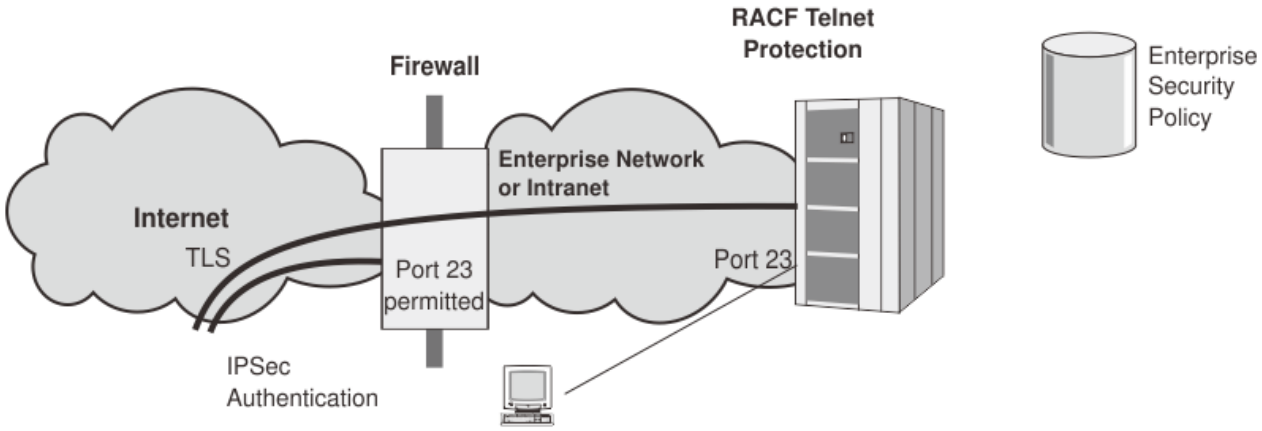


Figure 36. Secure and non-secure traffic using a single Telnet port

Express Logon Feature

With emulator products, the traditional method of authenticating the user is through user ID and password which is kept in sync with the host access control facility (RACF, ACF/2, AS/400 user management, etc.). The Express Logon Feature (ELF) simplifies user ID and password administration for users signing on to SNA applications using Telnet. ELF allows a user to use a TLS-authenticated X.509 certificate for authentication to the SNA application instead of using a user ID and password. ELF requires IBM Host Integration software. The Host Integration requirements depend on the configuration.

There are two network designs available; a two-tier or a three-tier approach. Both are discussed in Appendix C, “Express Logon Feature,” on page 1335.

TLS-enabled FTP

The Communications Server FTP server and client support Transport Layer Security (TLS). This support enables secure file transfer by providing data privacy, message authentication, and message integrity services for data sent and received using the FTP control and data connections.

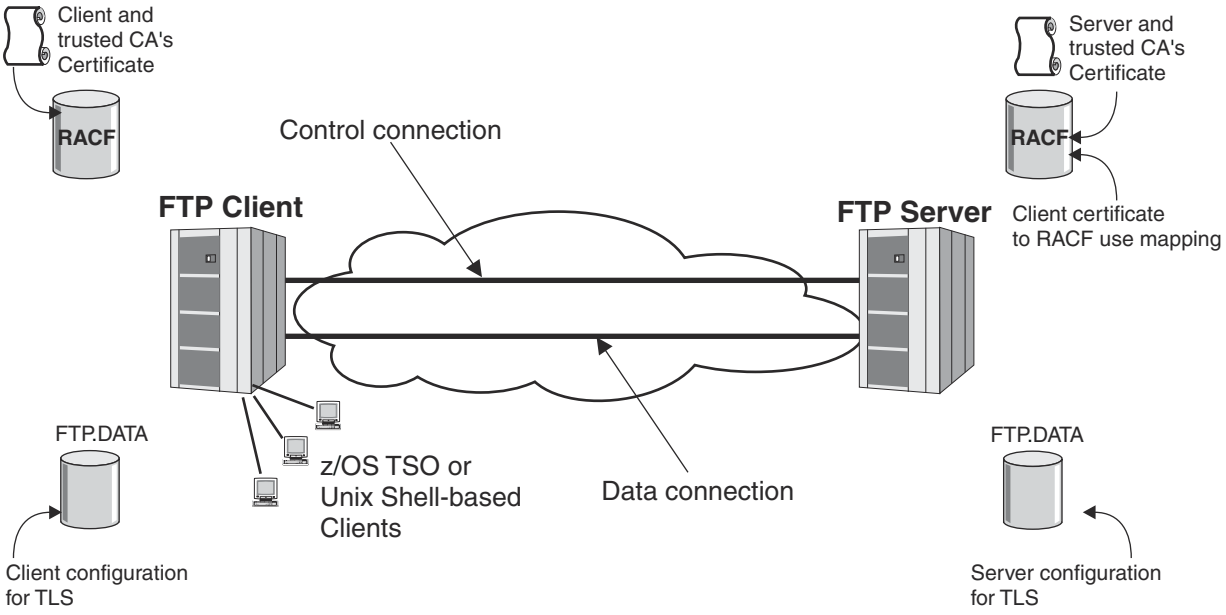


Figure 37. FTP client and server TLS overview

The TLS-enabled FTP server can be configured to run in two modes. Conditional mode allows an installation to use a single port for both TLS and non-TLS FTP control connections. In conditional mode, the FTP client and server negotiate the use of TLS based on a subset of the FTP security

negotiation functions documented in RFC 2228. After the use of TLS is negotiated, the TLS handshake is performed which establishes the TLS session and negotiates security parameters and session keys. Unconditional mode allows an installation to use a separate port for all TLS traffic. The port specified by the TLSPORT statement in FTP.DATA (port 990 by default) is the port designated for control connections for unconditional TLS mode. With unconditional mode, it is assumed that TLS is required, and after the FTP control connection is made, the TLS handshake is performed.

TLS secures the control connection and optionally the data connection. TLS for the data connection requires a TLS session for the control connection. FTP server configuration controls whether the FTP server requires TLS for the control and data connections. This TLS protection by connection type is negotiated during the FTP RFC 2228 negotiation that precedes the TLS handshake. During the lifetime of the control connection, the use of TLS or no TLS for the data connection can be requested by the FTP client using the FTP RFC 2228 commands.

FTP TLS optionally authenticates the client during the TLS handshake using a client X.509 certificate. FTP server configuration specifies whether TLS client authentication is required and what type of validation of the certificate is required. For example, the FTP server can be configured to map the client certificate to a RACF user ID and then verify that the user ID associated with the certificate matches the user ID entered by the user.

FTP TLS optionally simplifies the TLS handshake by resuming a previous TLS session instead of negotiating new security parameters and session keys. The FTP client and server configuration specifies whether reuse of a previous TLS session is required. For example, you can configure the FTP client to require session reuse of a previous TLS session and configure the FTP server to allow but not require session reuse.

Configuration to control TLS capabilities and options for both FTP client and server TLS are stored in the FTP.DATA data set.

Application Transparent Transport Layer Security

Communications Server provides for invocation of System SSL in the TCP transport layer of the stack on behalf of the TCP/IP application or middleware. Application Transparent Transport Layer Security (AT-TLS) support is controlled by the TTLS or NOTTLS parameter on the TCPCONFIG statement in the TCP/IP profile. When AT-TLS is enabled, AT-TLS policy statements processed by the Policy Agent define the security attributes for connections that match AT-TLS rules. This policy-driven support can be deployed transparently underneath many existing sockets, leaving the application unaware of the encryption and decryption being done on its behalf. Support is also provided for applications that need to negotiate TLS or need to participate in client authentication. These applications must be aware of AT-TLS support and use ioctl support provided by AT-TLS. AT-TLS supports the TLS, SSLv3, and SSLv2 protocols. For more details, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

Kerberos

Kerberos is a network authentication protocol that is designed to provide strong authentication for client/server applications using secret-key cryptography. The Kerberos network authentication protocol assumes that services and workstations communicate over an insecure network. It allows clients and servers to do either one way, or two way (mutual) authentication. It allows for data encryption and prevents passwords from having to be retyped to access networked services and also prevents their transmission in plain text over the network. This feature can help reduce the need to manage multiple passwords.

z/OS Integrated Security Services ships Kerberos version 5. You should write new applications to Kerberos version 5 and use z/OS Integrated Security Services.

The following Communications Server IP applications include support for the Kerberos version 5 security protocol:

- The UNIX System Services Telnet server allows clients supporting Kerberos version 5 (as described in RFC 1416) to log in to the shell environment using Kerberos as the authentication protocol.

- The FTP client and server support connections to or from other clients and servers supporting Kerberos version 5 authentication for the FTP protocol (as described in RFC 2228).
- The UNIX System Services RSH server can be configured to support client authentication using Kerberos from RSH clients supporting Kerberos version 5.

Secure Shell (SSH)

Secure Shell, or SSH is a client/server protocol that provides cryptographically secure remote login to a z/OS Unix shell, file transfers with sftp, and other capabilities over TCP connections.

z/OS OpenSSH, a port of open-source software OpenSSH, is a base element of z/OS.

OSPF authentication

Communications Server OSPF (Open Shortest Path First) dynamic routing protocol supports message authentication and message integrity of OSPF routing messages through the use of the OSPF MD5 Authentication security protocol as defined by RFC 2328. OSPF MD5 Authentication ensures that an unauthorized IP resource cannot inject OSPF routing messages into the network without detection, thus ensuring the integrity of the routing tables in the OSPF routing network.

OMPROUTE computes a secure MAC for the routing message using the MD5 algorithm. This MAC is sent with the routing message so that the message can be authenticated by the receiver.

SNMPv3

z/OS Communications Server SNMP supports SNMPv3. The legacy community-based protocols SNMPv1 and SNMPv2 are also supported. SNMPv3, defined in RFCs 3410 through 3415 is the standards-based solution for SNMP security. It is categorized as a User-based Security Model (USM) which provides different levels of security based on the user accessing the managed information. To support this security level, the SNMPv3 framework defines several security functions, such as USM for authentication and privacy, and view-based access control model (VACM) which provides the ability to limit access to different MIB objects on a per-user basis, and the use of authentication and data encryption for privacy. However, SNMP is not just enhanced security. It defines an architecture for SNMP management frameworks, with the intent that pieces of the architecture can advance over time without requiring the entire structure to be rewritten. For that reason, three major subsystems are defined:

- Message processing subsystem
- Security subsystem
- Access control subsystem

The framework is structured so that multiple models can be supported concurrently and replaced over time. For example, although there is a new message format for SNMPv3, messages created with the SNMPv1 and SNMPv2 formats can still be supported. Similarly, the user-based security model can be supported concurrently with the community-based security models previously used. For more information on SNMPv3 and configuring SNMPv3 support, see Chapter 23, “Simple Network Management Protocol,” on page 1237. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

Monitoring cryptographic network protection: z/OS encryption readiness technology (zERT)

z/OS Encryption Readiness Technology (zERT) is a Communications Server feature that provides information about the cryptographic network protection state of TCP and Enterprise Extender connections terminating on a z/OS system. zERT helps you answer the following questions:

- *What* TCP and Enterprise Extender traffic is being protected (and which is not)?
- *How* is that traffic being protected? For instance, what protocols are being used, which cryptographic algorithms are being used, and what key lengths?

- *Who* on my z/OS system is consuming or producing the network traffic, whether it is protected or not?
- *Where* is the remote endpoint for that traffic?

With zERT policy-based enforcement, you can write rules to enforce real-time compliance monitoring that can generate audit events and even take defensive actions based on the observed cryptographic protection attributes of each TCP connection.

z/OS Encryption Readiness Technology (zERT) Concepts

With the increasing number of corporate, industry, and government regulations regarding cryptographic protection of data in flight, as well as discoveries of weaknesses in existing cryptographic protocols and algorithms, it is important for z/OS administrators and auditors to be able to assess the quality of the cryptographic network protection being applied to their key z/OS workloads.

The landscape of cryptographic network protection on z/OS is varied and can be complex. Because of this, performing such an assessment can be very difficult. Consider the following:

- IBM provides two TLS/SSL protocol implementations on z/OS:
  - z/OS Cryptographic Services System SSL, which is available to software written in C or C++.
  - Java Secure Sockets Extension (JSSE), which available to Java software (and is written in Java itself).

  In addition, users have been known to port third party TLS implementations like OpenSSL onto z/OS. Programs and middleware that use System SSL directly or that use third party implementations usually have their own unique configuration methods to control the details of TLS/SSL protection. Programs and middleware that use JSSE often require a set of JSSE-specific environment variable or parameters to configure their TLS/SSL protection.
- z/OS Communications Server provides AT-TLS, which invokes System SSL on behalf of applications and middleware based on policies that you create. AT-TLS can protect traffic from any program that uses z/OS Unix System Services TCP sockets APIs, regardless of the programming language.
- z/OS Communications Server provides a full IPSec implementation, also configured through policies that you create.
- z/OS also supports the Secure Shell (SSH) protocol , although Communications Server does not directly use it. For more information of Secure Shell (SSH) protocol, see *z/OS OpenSSH* in *z/OS Introduction and Release Guide*.
- Each of these protocols allow the local and remote endpoints to negotiate the exact protection methods to be used for any given security session. This means understanding the z/OS configuration for a given application's cryptographic protection only tells you which attributes (protocol versions, algorithms, key lengths, and so on.) are possible - it usually does not tell you the exact set of protection attributes that were agreed upon for any specific security session that is established between the z/OS application and a remote endpoint.

With such variety of protocols, configuration methods, and audit and log records, it can be difficult to clearly understand the overall state of cryptographic network protection for your z/OS system.

zERT positions the z/OS TCP/IP stack as a central collection and reporting point for the cryptographic protection attributes for TLS, SSL, SSH, and IPSec security sessions that are protecting TCP and Enterprise Extender connections that terminate on the local stack. This collection and reporting function is called zERT discovery.

A second reporting function called zERT aggregation summarizes the repeated use of security sessions over a period of time (the system's SMF interval or the specified zERT aggregation INTVAL setting). By focusing more on the security session than on individual TCP or EE connections, zERT aggregation provides the same level of cryptographic detail as zERT discovery, but with significantly fewer SMF records. The zERT aggregation function requires the use of zERT discovery to collect the individual TCP and EE connection information that is then summarized by zERT aggregation.

IBM zERT Network Analyzer is a z/OS Management Facility (z/OSMF) task that provides a GUI-based tool for analyzing cryptographic protection characteristics of TCP and Enterprise Extender (EE) connections on your system, using SMF records generated by the zERT aggregation function.

With zERT policy-based enforcement (or simply "zERT enforcement"), the TCP/IP stack uses the cryptographic protection attributes observed by zERT discovery to enforce policy rules that you create based on your local network security requirements. zERT enforcement compares each TCP connection's observed cryptographic attributes to rules installed through the Policy Agent to take appropriate action when a connection’s protection attributes match those defined in a zERT enforcement rule. zERT enforcement rules and their associated actions are defined through the IBM® Configuration Assistant for z/OS Communications Server in z/OSMF. For acceptable protection, the default action is to silently allow the connection. For questionable or unacceptable protection, actions such as notification through messages, auditing through SMF records, and even dropping the connection can be configured.

What does zERT discovery manage and collect?

In order to understand zERT discovery processing and the information that zERT collects and records, it is important to understand the distinction between application connections and security sessions:

- An ***application connection*** is a TCP connection or Enterprise Extender (EE) connection (over UDP) over which two application programs communicate with each other.
- A ***security session*** is an instance of a secure path between two endpoints as defined by a cryptographic security protocol. Examples are TLS/SSL sessions, SSH sessions, and IPSec tunnels (as they apply to a given application connection).

Note: An application connection may have zero or more security sessions at any given time. Since "no protection" is a valid cryptographic protection state, zERT even reports on connections that do not have any recognized cryptographic protection. Likewise, a connection may be simultaneously protected by multiple security sessions, such as a TLS session and an IPSec tunnel. zERT collects and reports information for all of these scenarios.

How does zERT discovery collect the information?

zERT collects cryptographic protection attributes through two different mechanisms.

- For TCP and Enterprise Extender traffic, zERT can be notified of the attributes directly by ***cryptographic protocol providers (CPPs)*** that are specifically enabled for zERT. Such CPPs notifiy the TCP/IP stack when a significant event occurs regarding the security sessions they manage. z/OS IPSec, System SSL, IBM z/OS OpenSSH are all zERT-enabled CPPs.

  zERT-enabled CPPs generate these notifications when a security session is created, terminated, or renegotiated in a way that significantly changes the cryptographic protection attributes. In addition, IPSec generates a notification when an existing IPSec security association is assigned to protect a TCP or Enterprise Extender connection.
- For TCP traffic, zERT can learn the attributes through ***observational discovery***, which is the direct observation of the TCP stream as it passes through the TCP/IP stack for TLS, SSL and SSH handshake messages. This stream observation is performed only at very specific points in the life of the TCP connection for a very limited amount of time in order to avoid any negative impact to performance. The amount of observable information is limited. For detailed limitations of zERT discovery, see What are the limitations for zERT discovery? Despite its limitations, observational discovery can still provide basic information regarding cryptographic protection attributes for many TLS, SSL and SSH sessions that are managed by protocol providers that are not enabled for zERT.

How does zERT discovery provide the information?

As cryptographic protection attributes for TCP and Enterprise Extender connections are discovered and collected, you can direct zERT to write those attributes, along with relevant connection information, to the z/OS System Management Facility, to a network management application using the real-time NMI for zERT information, or to both. Regardless of the destination, the information is recorded using SMF 119 subtype 11 (zERT connection detail) records.

zERT connection detail records have the following general format:

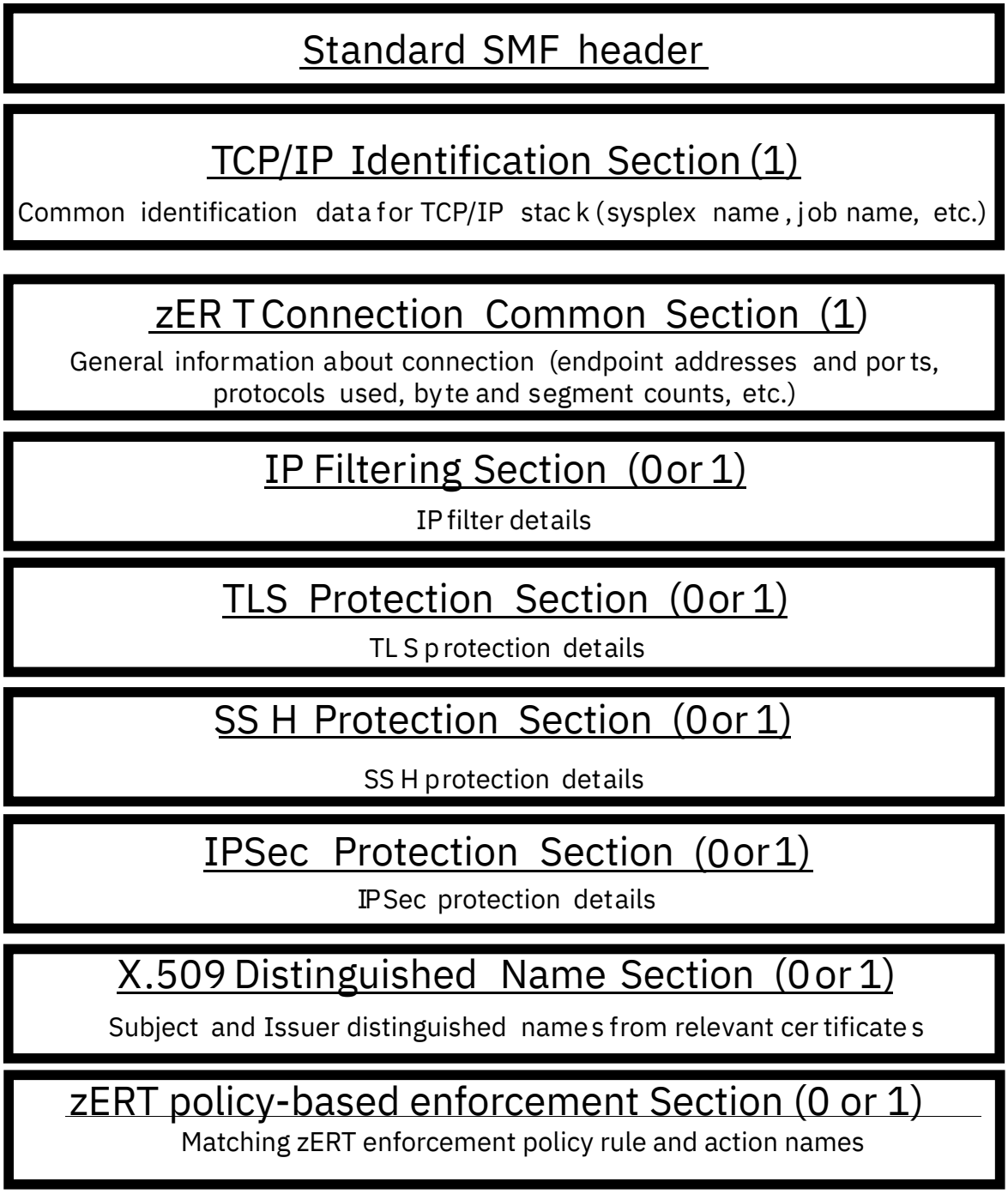


Figure 38. zERT connection detail (SMF 119 subtype 11) record layout

One or more zERT connection detail records are generated for each TCP and Enterprise Extender connection. The zERT Connection Common Section contains an event type field that indicates the reason each subtype 11 record is generated. The number of subtype 11 records written for a given connection depends on a couple of different factors:

- For TCP connections, the length of time that the connection exists can influence the number of records written.
  - TCP connections that last 10 seconds or longer will generate an SMF 119 subtype 11 record after the connection has existed for 10 seconds, and this is called a Connection Initiation event; when the TCP connection terminates, it will generate another record, and this is called a Connection Termination event.
  - TCP connections that last less than 10 seconds and do not experience a significant change in cryptographic protection during that time generate a single subtype 11 record when the TCP connection terminates. This is called a Short Connection Termination event.
- For EE connections, both a Connection Initiation and a Connection Termination event are generated.
- Any time there is a significant change in the cryptographic protection state after the initial cryptographic state is established, a subtype 11 record will be written to record the new state. This is called a Protection State Change event.
- If a TCP connection matches a zERT enforcement rule that requested an audit record, a subtype 11 record will be written. This is called a zERT Enforcement event.

  Restriction: zERT enforcement policy rules are not supported for EE connections.

Each connection-specific subtype 11 record contains information about the connection itself in the zERT Connection Common Section. This includes information like the local and remote connection endpoints, the attributes of the owner of the local socket such as userid and jobname, and information about which cryptographic protocols, if any, are being used to protect the connection.

For connections that have recognized TLS, SSL, SSH or IPSec attributes, the record contains a protocol-specific section that contains important cryptographic attributes of the security session that is protecting the connection. This includes the following attributes:

- Peer authentication method
- Cryptographic algorithms and key lengths in use for encryption, message authentication and key exchange
- Additional useful information like security session lifetime and X.509 certificate serial numbers

For connections protected by a security protocol that is using X.509 certificates for peer authentication, an X.509 Distinguished Name (DN) section is included with subject and issuer DNs from the end-entity certificates used for the connections.

Note: zERT does not store or record the values of secret keys, initialization vectors, or any other secret values that are negotiated or derived during cryptographic protocol handshakes.

For TCP connections that map to at least one zERT enforcement rule, the subtype 11 record contains a zERT policy-based enforcement section with the rules to which it was mapped.

Finally, if the connection is subject to an IP filter rule, the subtype 11 record also contains information about the inbound and outbound rules that are controlling the flow of traffic over the connection.

Tip:

1. Since Enterprise Extender (EE) traffic flows over UDP, the only cryptographic protocol that is capable of protecting this traffic is IPSec. As such you may see subtype 11 records for EE connections that have no recognized cryptographic protection, or that are protected by IPSec. In the IPSec case, each subtype 11 record will contain an IP filtering section and an IPSec section. In the case of no recognized protection, the subtype 11 records will have no protocol-specific section and may or may not have an IP filtering section, depending on the local IP filtering policy.

2. Since TCP traffic may be protected by not only IPSec, but also by TLS/SSL and SSH, subtype 11 records may have zero or more protocol-specific sections as described above. They may also have IP filtering sections in the same manner as Enterprise Extender traffic.

Subtype 11 records are also written in two special cases: when zERT discovery is enabled or disabled through the GLOBALCONFIG parameter.

For more information on the SMF 119 zERT connection detail (subtype 11) records, see zERT connection detail record (subtype 11) in z/OS Communications Server: IP Programmer's Guide and Reference.

What are the limitations for zERT discovery?

zERT discovery has the following limitations:

- zERT only monitors TCP and Enterprise Extender (EE) traffic. zERT does not monitor non-EE UDP traffic or any other IP protocol.
- zERT only monitors local traffic, meaning traffic that terminates at the local TCP/IP stack. zERT does not monitor traffic that is being routed through the local TCP/IP stack (this includes traffic that is generated within z/OS Container Extensions).
- zERT monitors for TLS, SSL, SSH and IPSec protection. No other cryptographic security protocols are supported.
- zERT recognition of TLSv1.3 protocols is only available on z/OS V2R4 or later.

The following z/OS cryptographic protocol providers are fully enabled for zERT:

- z/OS Communications Server IPSec
- z/OS Cryptographic Services System SSL (which also covers AT-TLS)
- z/OS OpenSSH
- IBM z/OS JSSE with the ZERTJSSE provider

Other TLS, SSL and SSH implementations running on z/OS are monitored using observational discovery, which provides a more limited set of security attributes. Specifically for the following situations:

- Only initial handshakes are observed. zERT has no visibility to early termination of TLS, SSL or SSH session or changes in security attributes made after the initial handshake is complete since the traffic on the connection is encrypted.
- The number of security attributes observed during the initial handshake is limited. Some attributes are not visible due to the use of encryption during cryptographic protocol negotiations, while others such as distinguished names and serial numbers from X.509 certificates are not observed in order to minimize stream observation's impact on performance.
- Significant portions of TLSv1.3 handshakes are encrypted, meaning zERT has no visibility to those handshake messages. This also means zERT can only infer if the handshake completed successfully.

There are a small number of cases where zERT is unable to monitor System SSL security sessions. These are cases where an application that calls System SSL directly begins a TLS/SSL handshakes in the middle of the application data stream instead of the very beginning of the stream. An API is provided for such applications to notify zERT that the handshake is about to occur. If the application uses that interface, then zERT will have visibility to the System SSL sessions. Otherwise, zERT will not have visibility to such sessions and will report them as having no recognized TLS/SSL protection. For more information about the API for notifying zERT, see SIOCSHSNOTIFY IOCTL in z/OS Communications Server: IP Programmer's Guide and Reference.

It is important to keep in mind that the number of SMF 119 subtype 11 records generated by the zERT discovery function could be very large, depending on the number of connections supported by your z/OS system as well as the frequency with which those connections are created and deleted. Here are some guidelines for understanding the possible volume of records:

- For TCP connections, at least one subtype 11 record will be written for every connection that lasts less than 10 seconds.

- For every TCP connection that lasts 10 seconds or longer, and for Enterprise Extender connections, at least two subtype 11 records will be written.
- The size of each subtype 11 record depends on the type and number of cryptographic security protocols being used to protect the subject connection. Remember that a single connection can be protected by zero or more protocols. All of the following are valid protection scenarios:
  - A connection is protected by a single TLS, SSL, SSH, or IPSec session.
  - A connection is protected by an IPsec tunnel as well as TLS, SSL or SSH.
  - Although it would be unusual, all three protocols could be applied to a single connection.
  - No cryptographic protection is applied (even in this case, the appropriate subtype 11 records will be written since it is a valid protection state).

If SMF 119 subtype 11 records are only needed for certain types of traffic or only in specific exception cases, you can use zERT policy-based enforcement to generate a targeted set of subtype 11 records.

A second zERT function called zERT aggregation can be used to report zERT information in a summarized format that requires far fewer SMF records than zERT discovery might generate.

What does zERT aggregation collect?

The zERT aggregation function creates summary records of the information collected by the zERT discovery function. The zERT aggregation function summarizes, or aggregates, the zERT discovery information in terms of the security sessions. These summary records are reported at regular intervals as SMF Type 119 subtype 12 records. The reporting interval is the global SMF interval unless a zERT aggregation-specific recording interval is configured with the INTVAL parameter on the GLOBALCONFIG ZERT AGGREGATION statement.

How does zERT aggregation summarize the information?

The zERT discovery information is aggregated into separate security session records according to the following rules:

- The connections must use the same server IP address and client IP address.
- The connections must use the same individual server port or server port range (see "How does zERT aggregation determine the server port?" on page 171 for more information).
- The job name associated with the local endpoint is the same for all connections.
- The local endpoint acts in the same server or client role for all connections. When the server and client for a TCP connection use the same IP address, separate session records are maintained for the server view and for the client view of the security session.
- The connections must use the same significant security cryptographic attributes for the security session being aggregated.
  - **Significant security cryptographic attributes** are characteristics of the security session, such as cipher, authentication method, etc., that make a substantial difference to the strength of the security coverage being provided by the security session.
  - The zERT discovery function also collects and records **informational security cryptographic attributes**, such as tunnel ID, handshake time limits, etc. These attributes are of interest for a given TCP connection but do not have any impact on the protection strength provided. The zERT aggregation function does not report or consider informational security cryptographic attributes when aggregating connection information.

Consider the following example:

- Client-A at 10.11.1.2 establishes a series of six TCP connections to server at 10.12.3.3 using port 50. All connections use the same TLS security coverage characteristics.
- Client-B at 10.11.1.4 establishes a series of three TCP connections to the same server at 10.12.3.3, port 50. All connections use both IPSec and TLS security coverage.

- Client-C at 10.11.1.4 also establishes three connections to the same server, and all connections use the same IPSec tunnels and TLS sessions as the Client-B connections.

The zERT discovery function would generate at least 12 SMF records for this example if all TCP connections were all short-lived, meaning less than 10 seconds in duration, or at least 24 records if all TCP connections were long-lived (more than 10 seconds in duration).

The zERT aggregation function would instead generate just three SMF records:

- The connections involving Client-A would be aggregated into a single SMF record.
- The connections involving the other clients would be aggregated into two SMF records:
  - One SMF record would report the use of an IPSec tunnel for the connections involving both Client-B and Client-C.
  - One SMF record would report the use of the TLS session for the connections involving both Client-B and Client-C.

How does zERT aggregation provide the information?

You can direct zERT aggregation to write the summarized records to the z/OS System Management Facility, to a network management application using the real-time NMI for zERT summary information, or to both. Regardless of the destination, the information is recorded using SMF 119 subtype 12 (zERT summary) records.

zERT summary records have the following general format:

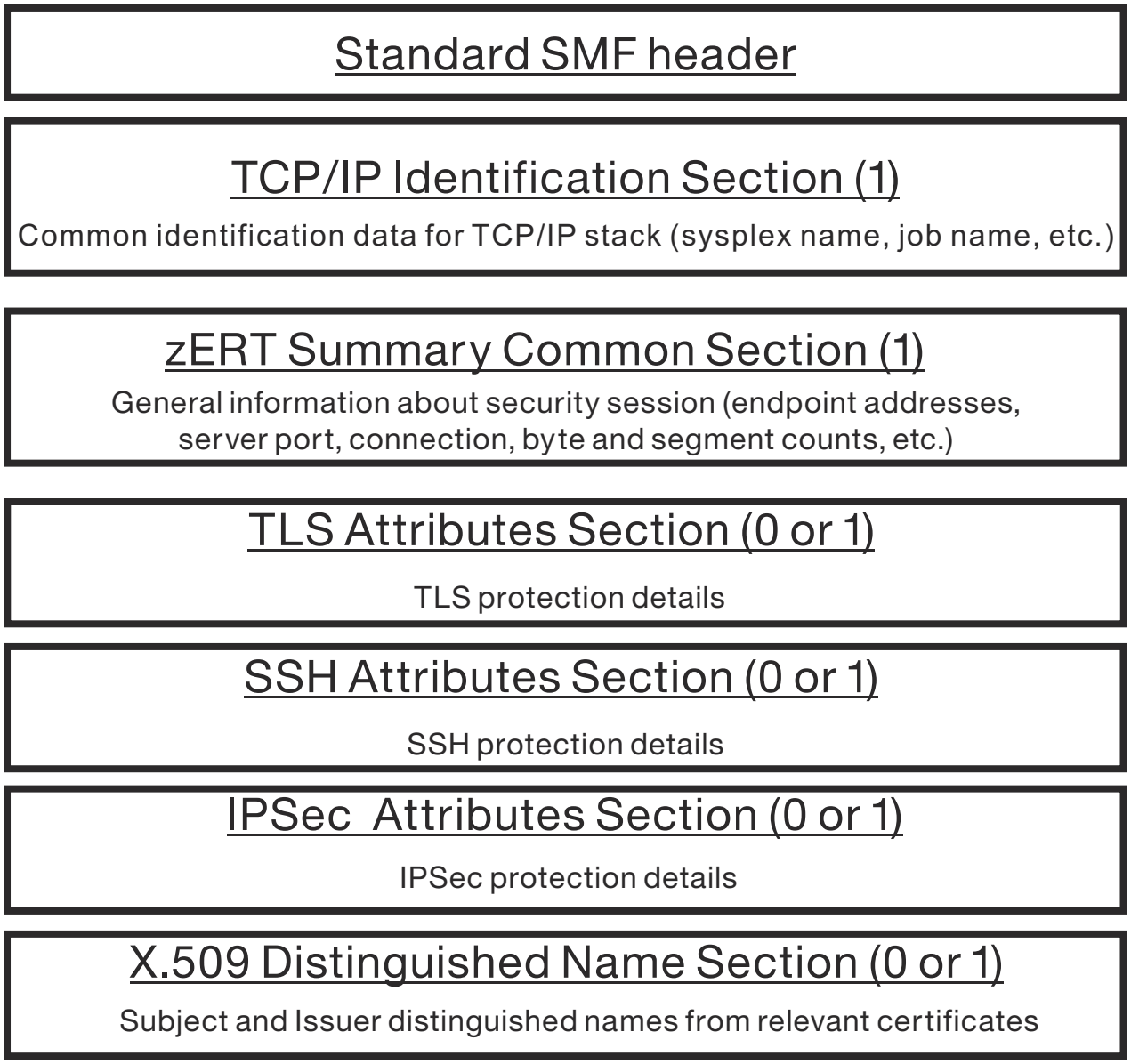


Figure 39. zERT summary (SMF 119 subtype 12) record layout

One zERT summary record is generated at the end of each SMF/INTVAL interval for each unique security session that was used during that interval. Each zERT summary record contains information about the security session in the zERT Summary Common Section. This includes the server and client security session endpoints, the server port or port range, attributes of the owner of the local socket such as userid and jobname, and statistical information about the usage of the security session. This statistical information includes the following information for the security session:

- The total number of TCP or EE connections that were ever covered by this session.
- The number of active TCP or EE connections that are covered by this session at the time of reporting.
- The number of partial TCP or EE connections that were ever covered by this session. A *partial TCP connection* is one that changed significant security attributes at some point, for instance changed from having no recognized cryptographic protection to having a recognized cryptographic protection.
- The number of short lived TCP connections that were covered by this session. EE connections are never considered to be short lived connections.
- The number of inbound and outbound data bytes and segments processed by the TCP or EE connections when they were covered by this security session.

The record contains both the values for these statistics at the end of the current reporting period as well as the values at the start of the current reporting period. You can subtract the ending values from the starting values to determine the activity for the SMF interval.

Security sessions that provide no recognized cryptographic protection, meaning they represent connections flowing as clear text or are protected in a way that zERT discovery does not recognize, will only contain the TCP/IP Identification and zERT Summary Common sections. For security sessions that represent TLS, SSL, SSH or IPSec attributes, the record contains a protocol-specific section that records the significant cryptographic attributes of the security session.

Guideline: A zERT summary SMF record will have at most one protocol-specific section since it represents a single security session rather than individual connections.

If X.509 Distinguished Names are used by the security session, the record also includes that section.

Subtype 12 records are also written in two special cases: when zERT aggregation is enabled or disabled through the GLOBALCONFIG parameter.

For more information on the SMF 119 zERT summary (subtype 12) records, see zERT Summary record (subtype 12) in z/OS Communications Server: IP Programmer's Guide and Reference.

How does zERT aggregation determine the server port?

For most server applications, determining the server port is simple: it is the port that the server is listening on for new TCP connection requests. For FTP processing, however, there are some complications to that simple scheme.

- For IPv4 *active* data connections, meaning those that are not firewall-friendly, the connections are initiated outbound from the FTP server to the FTP client. In this case, even though outbound connections are typically considered to indicate a client connecting outbound to a server, the zERT aggregation function recognizes these outbound connections as server-side connections. For typical FTP servers that listen on port 21, these data connections are aggregated using server-side port 20.
- For IPv4 *passive* data connections, meaning those that are firewall-friendly, or for all IPv6 data connections, the connections are initiated outbound from the FTP client to the FTP server using an ephemeral server port that was previously provided to the client. The z/OS FTP server can be configured, using the PASSIVEDATAPORTS parameter, to limit the ephemeral server port value to a specific range of values. When PASSIVEDATAPORTS is in use:
  - At the FTP server, all inbound data connections using a port within the configured port range are aggregated into a single aggregation record. When the aggregation record is written as an SMF record, the SMF record reports a starting and ending server port range that matches the active PASSIVEDATAPORTS definition.

    Results: If multiple FTP servers (for instance, ports 21 and 621) operate at the same server IP address, and one or more FTP servers use the same PASSIVEDATAPORTS range for their data connections, the data connections to the FTP servers with the same PASSIVEDATAPORTS range are aggregated into the same zERT summary SMF record, assuming all other aggregation criteria is the same for the connections. If you want separate zERT summary records for the data connections for the different FTP servers, specify a different PASSIVEDATAPORTS range for each individual FTP server.
  - At the FTP client, the use of PASSIVEDATAPORTS by the server to assign the ephemeral port is unknown, so all FTP data connections are aggregated into a single aggregation record. When the aggregation record is written as an SMF record, the SMF record reports a starting and ending server port range equal to the entire ephemeral port range (1024-65535).

    Results: If clients connect to multiple FTP servers (for instance, ports 21 and 621) at the same server IP address, all the data connections to the multiple FTP servers are aggregated into the same zERT summary SMF record, assuming all other aggregation criteria is the same for the connections.

How does zERT enforcement work?

zERT enforcement depends upon the data collected by zERT discovery. zERT enforcement policy rules describe acceptable or unacceptable cryptographic protection attributes for different types of z/OS connections and any appropriate actions to take when a connection matches the rule. You create zERT enforcement rules and actions through the IBM® Configuration Assistant for z/OS Communications Server in z/OSMF. These rules should reflect your local network encryption standards and requirements. Once the rules and actions are defined, the NCA creates a policy file that is stored on the z/OS system and processed by the Policy Agent. The Policy Agent processing installs the zERT enforcement rules into the TCP/IP stack which then compares each new TCP connection that terminates on that stack against the set of zERT enforcement rules. If the connection matches a rule, the actions associated with that rule are taken. The following actions are supported:

- Log a message through syslogd
- Log a message to the console (TCP/IP job log)
- Write an audit record (SMF zERT Connection Detail (type 119, subtype 11) with an event type 7 which indicates that it was written by zERT Enforcement)
- Reset the connection

These actions can be used in combination on a single rule. The default action is to silently allow the connection.

A connection is evaluated against zERT enforcement policy rules based on the security protocol(s) used to protect it. If a connection is protected by multiple protocols, it could match to more than one rule.

For example: when a connection is protected by IPsec and TLS, the connection can map to a zERT enforcement rule for IPsec and a zERT enforcement rule for TLS. If both the rules request an audit action, then two zERT Enforcement records will be written for the connection.

zERT enforcement can act upon the data collected by zERT observational discovery and from the IPsec function within the TCP/IP stack.

Restriction: zERT enforcement policy rules are not supported for Enterprise Extender (EE) connections.

Using z/OS Encryption Readiness Technology

Separate parameters in the TCP/IP profile control whether zERT discovery or aggregation is enabled and where the SMF 119 subtype 11 or 12 records should be written. If you want to collect zERT information, it is important to remember to not only enable zERT discovery or aggregation, but also to enable recording to one or both of SMF and real-time NMI, since each of these have separate controls. To use zERT policy-based enforcement, zERT discovery must be enabled and the Policy Agent must be configured to process and install the zERT enforcement policy you create through the IBM® Configuration Assistant for z/OS Communications Server in z/OSMF.

Enabling zERT discovery

You turn the zERT discovery function on in the TCP/IP stack by specifying the GLOBALCONFIG ZERT parameter in the TCP/IP profile data set. Enabling zERT causes the TCP/IP stack to record and collect cryptographic protection information for new TCP and Enterprise Extender connections that terminate at the stack.

Guidelines:

- Enabling zERT discovery is required for zERT aggregation, but does not automatically enable zERT aggregation. See “Enabling zERT aggregation” on page 173 for details.
- Enabling zERT discovery is required to use zERT policy-based enforcement but there are additional requirements. See “Enabling zERT policy-based enforcement” on page 173 for details.

Results: If you dynamically enable zERT discovery using VARY OBEYFILE, zERT does not collect information about TCP and Enterprise Extender connections that were established before enabling zERT.

For more information on enabling zERT, see GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Enabling zERT aggregation

You turn the zERT aggregation function on in the TCP/IP stack by specifying the GLOBALCONFIG ZERT AGGREGATION parameter in the TCP/IP profile data set. Enabling zERT aggregation causes the TCP/IP stack to collect cryptographic protection summary information for security sessions involving TCP and Enterprise Extender connections that terminate at the stack.

Guideline: Because zERT aggregation summarizes information collected by zERT discovery processing, zERT discovery must be enabled for zERT aggregation to operate; however, you do not have to enable SMF recording of zERT connection detail records for zERT aggregation to function correctly.

Results: If you dynamically enable zERT aggregation using VARY OBEYFILE, zERT does not collect summary information about TCP and Enterprise Extender connections that were established before enabling zERT aggregation.

Defining the zERT aggregation recording interval

By default, zERT aggregation writes the data it has collected in the form of SMF 119 subtype 12 (zERT summary) records on each SMF interval. If you want to reduce the frequency at which these records are written, you can specify a zERT aggregation-specific recording interval between 1 and 24 hours (in one hour increments) by specifying the INTVAL sub-parameter of the GLOBALCONFIG ZERT AGGregation parameter in the TCP/IP profile data set.

If you specify an INTVAL value, it is recommended that the value divides evenly into 24 hours to ensure the SMF 119-12 records are written at the same times each day.

To specify exactly when the INTVAL intervals are to begin, you can specify the SYNCVAL sub-parameter of INTVAL. SYNCVAL defines the reference time from which the INTVAL intervals begin to be applied. SYNCVAL is specified as a time of day in 24 hour clock format (hh:mm). For example, if SYNCVAL is specified as 06:00 (6 AM) and INTVAL is set to 12 hours, then zERT aggregation will write its SMF records at 6 AM and 6 PM daily. If not specified, the default SYNCVAL is midnight (00:00).

For more information on enabling zERT aggregation, see GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Decreasing the frequency at which zERT summary records are written may increase the amount of 64-bit pageable, private memory needed, because the zERT aggregation information is held longer in memory before being written out to SMF. Any increase in memory usage will be dependent on the nature of network connections as well as the length of the recording interval. Environments where the majority of the connection patterns are consistent throughout a 24 hour period will require less memory than those where connection patterns vary significantly throughout the day. A longer recording interval increases the likelihood of greater memory consumption. In environments where 64-bit private storage is limited, consider configuring a smaller zERT aggregation recording interval in order to allow for more frequent flushing of zERT Aggregation information.

For more information on displaying zERT aggregation storage information, see DISPLAY TCPIP,,STOR in z/OS Communications Server: IP System Administrator's Commands.

Enabling zERT policy-based enforcement

A few steps are required to enable zERT policy-based enforcement:

1. Since zERT enforcement depends on data that zERT discovery collects for each TCP connection, the first step is to enable zERT discovery as described in “Enabling zERT discovery” on page 172.
2. Configure the Policy Agent with the name of the z/OS Unix file or z/OS dataset that contains the zERT enforcement policy rules and actions. Specify the policy file path or data set name on the ZERTConfig statement in the image configuration file for the affected TCP/IP stack.

3. If you plan to use the audit action on any of your zERT enforcement rules, specify the appropriate by `policy` parameters on the SMFCONFIG and/or NETMONITOR statements in the TCP/IP profile data set. See "Selecting a destination for zERT discovery SMF records" on page 174 for more information.
4. Use the IBM® Configuration Assistant for z/OS Communications Server in z/OSMF to create your zERT enforcement policy rules and save the generated zERT enforcement policy file in the file or data set specified in step 2 above.
5. Start or refresh the Policy Agent to process and install your zERT enforcement policy rules.

Selecting a destination for zERT discovery SMF records

If you enable zERT discovery, then you can decide where you want the collected data to be recorded as SMF 119 subtype 11 records. By default, the SMF records are not written anywhere, so you must choose one of the following destinations in order to get the SMF records.

- If you want the zERT connection detail (SMF 119 subtype 11) records to be written to the z/OS System Management Facility for connection initiation, termination and when security characteristics for the connection change (event types 1-4), specify the SMFCONFIG TYPE119 ZERTDETAIL parameter in the TCP/IP profile data set. Note that you must also have the recording of SMF 119 records specified in your SMF parmlib member.
- If you want the policy-driven zERT connection detail records (SMF 119 subtype 11, event type 7) to be written to the z/OS® System Management Facility, specify the SMFCONFIG TYPE119 ZERTDETAILBYPOLICY parameter in the TCP/IP profile data set.

  Note: You must also have the recording of SMF 119 records specified in your SMF parmlib member.
- If you use a network management application that consumes zERT connection detail SMF records for connection initiation, termination, and when security characteristics for the connection change (event types 1-4) through the real-time zERT Detail NMI service (SYSTCPER), specify the NETMONITOR ZERTSERVICE parameter in the TCP/IP profile data set. When this service is active, network management applications with the necessary SAF permission can connect to the service to receive zERT connection detail SMF records in near real-time.
- If you use a network management application that consumes policy-driven zERT connection detail SMF records through the real-time zERT Detail NMI service (SYSTCPER), specify the NETMONITOR ZERTSERVICEBYPOLICY parameter in the TCP/IP profile data set. When this service is active, network management applications with the necessary SAF permission can connect to the service to receive zERT connection detail SMF records in near real-time.

  Note: If GLOBALCONFIG ZERT is not configured, then no zERT information will be collected, so even if you configure a recording destination, no SMF 119 subtype 11 records are generated.

For more information on configuring SMFCONFIG or NETMONITOR, see SMFCONFIG statement and NETMONITOR statement in z/OS Communications Server: IP Configuration Reference.

Selecting a destination for zERT aggregation SMF records

If you enable zERT aggregation, then you can decide where you want the collected data to be written as SMF 119 subtype 12 records. By default, the SMF records are not written anywhere, so you must choose one of the following destinations in order to get the SMF records.

- If you want the zERT summary (SMF 119 subtype 12) records to be written to the z/OS System Management Facility, specify the SMFCONFIG ZERTSUMMARY parameter in the TCP/IP profile data set.

  Note: You must also enable the recording of SMF 119 records, and enable interval record processing, in your SMF parmlib member.

  Note: If you want to use IBM zERT Network Analyzer, you must specify the SMFCONFIG ZERTSUMMARY parameter in the TCP/IP profile data set.
- If you use a network management application that consumes zERT summary SMF records through the real-time zERT Summary NMI service, specify the NETMONITOR ZERTSUMMARY parameter in the TCP/IP profile data set. When this service is active, network management applications with the

necessary SAF permission can connect to the service to receive zERT summary SMF records in near real-time.

Note: You must also enable interval record processing in your SMF parmlib member.

Note: If GLOBALCONFIG ZERT AGGREGATION is not configured, then no zERT aggregation information will be collected, so even if you configure a recording destination, no SMF 119 subtype 12 records are generated.

Results: If you dynamically change your destination selection by specifying SMFCONFIG NOZERTSUMMARY, a final set of zERT summary SMF records are written to the z/OS System Management Facility. This final set is not written to the real-time zERT Summary NMI service if that has also been selected as a destination for the zERT summary records.

For more information on configuring SMFCONFIG or NETMONITOR, see SMFCONFIG statement and NETMONITOR statement in z/OS Communications Server: IP Configuration Reference.

Disabling zERT discovery

You turn the zERT discovery function off in the TCP/IP stack by specifying the GLOBALCONFIG NOZERT parameter in the TCP/IP profile data set.

Guideline: zERT discovery is disabled by default.

Results:

- Disabling zERT discovery also disables zERT aggregation.
- Disabling zERT discovery also disables the data collection upon which zERT policy-based enforcement depends.

Disabling zERT discovery causes the TCP/IP stack to not monitor the cryptographic characteristics of TCP and Enterprise Extender connections that terminate at the stack.

Results: If you dynamically disable zERT discovery using VARY OBEYFILE, any previous information collected by zERT discovery and aggregation is discarded.

For more information on disabling zERT, see GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Disabling zERT aggregation

You turn off the zERT aggregation function in the TCP/IP stack by either specifying either the GLOBALCONFIG ZERT NOAGGREGATION parameter or the GLOBALCONFIG NOZERT parameter in the TCP/IP profile data set.

Guideline: This is the default behavior.

Results: Using GLOBALCONFIG ZERT NOAGGREGATION to disable zERT aggregation has no effect on zERT discovery processing or zERT enforcement.

Disabling zERT aggregation causes the TCP/IP stack to not maintain summary statistics of the cryptographic characteristics of TCP and Enterprise Extender security sessions that terminate at the stack. Before disabling the zERT aggregation function, a final set of SMF Type 119 subtype 12 records is generated, if recording of zERT summary records is active. The final set of records is written to the z/OS System Management Facility (SMF), the real-time zERT Summary NMI service, or both, depending on your choice of destination at the time zERT aggregation is disabled.

Results: If you dynamically disable zERT aggregation or zERT discovery using VARY OBEYFILE, any previous information collected by zERT aggregation is discarded.

For more information on disabling zERT aggregation, see GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Using IBM zERT Network Analyzer

z/OS Management Facility (z/OSMF) provides a web browser interface for a variety of z/OS system management functions. IBM zERT Network Analyzer runs as a z/OSMF task and provides a structured interface for analyzing cryptographic protection data about your TCP and Enterprise Extender (EE) traffic that is collected in the form of zERT summary (Type 119 subtype 12) SMF records.

You can use the IBM zERT Network Analyzer task to perform the following functions:

- Import one of more SMF dump data sets into the IBM zERT Network Analyzer database. As the SMF data set is imported, IBM zERT Network Analyzer extracts information from the zERT Summary SMF records in the data sets and organizes the information into representations of security sessions reported in the SMF records.
  - If the SMF records are written to an SMF data set, the records must be dumped to the data set using the IFASMFDP program.
  - If the SMF records are written to an SMF log stream, the records must be dumped to the data stream using the IFASMFDL program.

  Note: SMF data sets that you intend to import into IBM zERT Network Analyzer must be a cataloged sequential file.
- Create and run queries to filter the imported security session data by date, system, endpoint or cryptographic protection attributes.
- Examine the results of your query.
  - The summary view provides an overall view of security sessions organized by endpoint role of the local z/OS system: TCP server, TCP client, or Enterprise Extender (EE) peer.
  - Expand each summary row to examine details about the foreign endpoints that established connections with the local endpoint in the summary row.
  - The security session view provides the cryptographic details about the security sessions that are used by the application connections.
- Export your query results to a comma-separated values (CSV) format file for additional analysis using a spreadsheet or analysis tool of your choice.

For additional information about using IBM zERT Network Analyzer in z/OSMF, see IBM z/OS Management Facility Configuration Guide.

Defining a zERT enforcement policy

Options for configuring a zERT enforcement policy

zERT enforcement is configured using a set of configuration statements and parameters coded in a file, which is parsed by the Policy Agent to establish the zERT enforcement policy for each TCP/IP stack. There are two alternatives for creating the Policy Agent files.

- "Option 1: Use the IBM Configuration Assistant for z/OS Communications Server" on page 176
- "Option 2: Manual configuration" on page 177
- The policy agent configuration must also be considered when configuring zERT enforcement policy. See "Specifying the zERT enforcement configuration file based on Policy Agent role" on page 177.

Option 1: Use the IBM Configuration Assistant for z/OS Communications Server

The IBM® Configuration Assistant for z/OS® Communications Server (Network Configuration Assistant), an optional GUI-based tool, provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the Network Configuration Assistant to generate the Policy Agent files.

The Network Configuration Assistant is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the Network Configuration Assistant in z/OSMF, the Network Configuration Assistant runs natively in the z/OS system and you can access it through a web browser.

Through a series of wizards and online help panels, you can use the Network Configuration Assistant to create zERT policy configuration files for any number of z/OS images with any number of TCP/IP stacks per image. The Network Configuration Assistant provides the following reusable objects for the zERT enforcement technology:

- Traffic descriptors that define a traffic pattern or application, by describing the TCP traffic with ports, jobname, userid, and the direction the connection is established.
- Protection characteristics that describe how the traffic is protected including the symmetric encryption, message authentication, and key exchange algorithms.
- Address groups that define one or more IP addresses or subnets.
- Reusable rules that link the characteristics of an application (traffic descriptor and IP addresses), protection characteristics, and the actions to take.
- Reusable rule sets group together all rules for a particular security protocol. There can be a rule set for the TLS/SSL protocol, for the IPsec protocol, and for the SSH protocol. There can also be a rule set for traffic that has no recognized protection, which will be referred to as having a "None" security protocol.

For each TCP/IP stack, you can include up to four reusable rule sets, one for each security protocol – TLS/SSL, IPsec, SSH, and None. If a rule set is included for a security protocol, then the rule set will be searched for each TCP connection protected by that protocol.

The Network Configuration Assistant comes with a number of IBM-supplied traffic descriptors that are easily applied. You can also create your own traffic descriptors.

The Network Configuration Assistant can dramatically reduce the amount of time that is required to create zERT enforcement policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS security, using the GUI can help you ensure that you have a consistent and easily manageable interface for implementing zERT policy-based enforcement.

Option 2: Manual configuration

You can manually create the zERT enforcement policy configuration files by coding all the required statements in a z/OS® UNIX file or MVS™ data set. zERT enforcement rules and zERT enforcement actions provide guidance on manually configuring the zERT enforcement policy, along with general tips that are helpful regardless of the configuration option being used. For details about the zERT policy statements, see z/OS Communications Server: IP Configuration Reference.

Specifying the zERT enforcement configuration file based on Policy Agent role

The Policy Agent can act as a policy server, a policy client, or neither. For more information on these different roles, see “Policy types and infrastructure overview” on page 787. Regardless of which option is used to configure zERT enforcement policies, the resulting configuration files need to be specified using different statements, depending on the role of the Policy Agent.

- If you are not using a policy client/policy server environment, specify the configuration files using the ZERTConfig statement on the single Policy Agent.
- If you are using the Policy Agent as a policy client that retrieves zERT enforcement policies from the policy server, specify the configuration files using the DynamicConfigPolicyLoad statement on the policy server.
- If you are using the Policy Agent as a policy client, but the policy client does not retrieve zERT enforcement policies from the policy server, specify the configuration files using the ZERTConfig statement on the policy client.

When information refers to configuration files, keep in mind where the files should exist, based on the role of the Policy Agent.

Configuring a zERT enforcement policy

zERT enforcement policies are processed by Policy Agent and installed into a z/OS® Communications Server TCP/IP stack.

Within the zERT enforcement policy file, zERT rules are defined for different security protocols:

- TLS/SSL
- IPSec
- SSH

zERT enforcement rules can also be defined for connections that have no recognized cryptographic protection. This will be referred to as security protocol None.

It is helpful to design your zERT enforcement rules by security protocol. For example, if you use both TLS and IPsec to protect TCP traffic, you can create a group of rules that describe your TLS/SSL requirements and a group of rules that that describe your IPsec requirements. This grouping of rules by security protocol, or "rule sets", reflect the way the TCP/IP stack evaluates zERT enforcement rules. When a specific type of cryptographic protection is detected for a TCP connection, or when it is determined that there is no recognized cryptographic protection for a TCP connection, the TCP/IP stack evaluates the connection and its protection attributes against rules in the appropriate rule set.

The grouping of rules into a rule set in the policy agent file is based on the setting of the SecurityProtocol parameter on the ZERTRule. You can configure one or more of the following types of rule sets:

- TLS rule set (zERT enforcement rules with a SecurityProtocol of TLS) - Rules within this rule set are searched when zERT discovery detects that a TCP connection is protected by TLS or SSL.
- IPsec rule set (zERT enforcement rules with a SecurityProtocol of IPsec) - Rules within this rule set are searched when zERT discovery detects that a TCP connection is protected by IPsec.
- SSH rule set (zERT enforcement rules with a SecurityProtocol of SSH) - Rules within this rule set are searched when zERT discovery detects that a TCP connection is protected by SSH.
- No recognized protection rule set (zERT enforcement rules with a SecurityProtocol of None) - Rules within this rule set are searched when zERT discovery determines that a TCP connection has no recognized cryptographic protection.

An effective approach for creating a rule set is to use a model where you have a general rule that covers the acceptable levels of protection for the security protocol, followed by one or more specific rules that define exceptions to the general rule. These rules can either define an exception that is acceptable or one that requires notification for further investigation. Finally, you can have a catch-all rule. Any connection that does not match the general rule or the specific rules will match the catch-all rule.

For example, the general rule in a TLS rule set might indicate that TLSv1.2 and TLSv1.3 are the acceptable TLS versions. This general rule might also include various acceptable algorithms. A specific rule might allow clients connecting to a specific application from a defined subnet to use TLSv1.1. A second specific rule might ban all use of SSLv2 and SSLv3. Finally, a catch-all rule might audit any connections using TLS that do not match the higher priority rules in the TLS/SSL rule set.

Tip: The IBM Configuration Assistant for z/OS Communications Server helps you create and organize rule sets, as well as, configuring general, specific, and catch-all rules.

zERT enforcement rules

This topic provides guidance on manually configuring a zERT enforcement rule, along with general tips that are helpful for both manual configuration and configuration with the Network Configuration Assistant. For details about zERT enforcement policy statements, see zERT policy statement in z/OS Communications Server: IP Configuration Reference.

A ZERTRule statement consists of a set of conditions that are compared against a TCP connection. When a connection's attributes match the conditions specified in a rule, zERT enforces the actions configured on the ZERTAction referenced by the rule. The rule conditions must include SecurityProtocol with one of the following values - TLS, IPsec, SSH, or None.

A ZERTRule with SecurityProtocol TLS can include the following additional conditions which describe the TLS cryptographic protection characteristics for a connection that matches the rule:

- ZERTTLSProtocol – one or more TLS protocol versions, such as TLSv1.2
- ZERTSymmetricEncryption – one or more symmetric encryption algorithms, such as AES_GCM_128

- ZERTMessageAuthentication – one or more message authentication algorithms, such as HMAC_SHA2_256
- ZERTKeyExchange – one or more key exchange algorithms, such as ECDHE_RSA

A ZERTRule with SecurityProtocol SSH can include the following additional conditions which describe the SSH cryptographic protection characteristics for a connection that matches the rule:

- ZERTSSHProtocol – one or more SSH protocol versions, such as SSHv2
- ZERTSymmetricEncryption – one or more symmetric encryption algorithms, such as AES_CTR_256
- ZERTMessageAuthentication – one or more message authentication algorithms, such as HMAC_SHA2_256
- ZERTKeyExchange – one or more key exchange algorithms, such as Curve_25519_SHA256

A ZERTRule with SecurityProtocol IPsec can include the following additional conditions which describe the IPsec cryptographic protection characteristics for a connection that matches the rule. Note that these are the characteristics negotiated for a phase 2 IPsec tunnel which protects the connection data:

- ZERTSymmetricEncryption – one or more symmetric encryption algorithms, such as AES_CBC_256
- ZERTMessageAuthentication – one or more message authentication algorithms, such as HMAC_SHA2_384_192

A rule with SecurityProtocol None has no recognized cryptographic protection, so none of the cryptographic protection conditions are allowed.

All ZERT rules, regardless of the SecurityProtocol setting can also include the following conditions which describe the traffic characteristics of a connection that matches the rule:

- LocalAddr - Local IP address or addresses
- RemoteAddr - Remote IP address or addresses
- Protocol – Transport protocol (TCP is the only supported value)
- LocalPortRange - Local port or ports
- RemotePortRange - Remote port or ports
- Jobname - Job name of the local z/OS application
- Userid – z/OS user ID that opened the local socket
- TcpConnectionDirection – Direction in which the TCP connection is initiated

The LocalAddr and RemoteAddr can be specified directly on the ZERTRule. The other conditions are parameters on a ConnectionDescriptor statement.

Each ZERTRule statement also has a priority which determines the order in which the rules are searched within a rule set. Priority values can be integers in the range 1 - 2000000000, with 2000000000 being the highest priority. When assigning priorities, you should skip some values to allow for future rule insertion between existing rules.

Finally, the ZERTRule references a ZERTAction which indicates the actions that should be taken for connections that match the ZERTRule.

If you are using a model where you have a general rule to cover the acceptable levels of protection for a security protocol, the rule would typically indicate the cryptographic protection characteristics but not include any traffic characteristic conditions. Specific rules that define exceptions to the general rule would use a combination of traffic characteristics and cryptographic protection characteristics to describe the connections that map to the rule. A catch-all rule, intended to match any connection that does not match the general or specific rules, would include only the SecurityProtocol value with no traffic characteristic conditions and no protection conditions.

zERT enforcement actions

This topic provides guidance on manually configuring a zERT enforcement rule, along with general tips that are helpful for both manual configuration and configuration with the Network Configuration

Assistant. For details about zERT enforcement policy statements, see zERT policy statement in z/OS Communications Server: IP Configuration Reference.

The ZERTAction provides the following actions that can be taken when a TCP connection matches a zERT enforcement rule.

- Write an audit record to SMF and/or the NMI SYSTCPER service. The audit record is an SMF zERT Connection Detail record (type 119, subtype 11) with an event type 7. This action is configured with the AuditRecord parameter and is off by default.

  Restriction:

  – SMFCONFIG TYPE119 ZERTDETAILBYPOLICY must be configured in the TCP/IP profile for SMF records to be generated.

  – NETMONITOR ZERTSERVICEBYPOLICY must be configured in the TCP/IP profile for the records to be written to the SYSTCPER NMI service.

- Log a message through syslogd. This action is configured with the LogSyslogd parameter and is off by default. If ResetTCPConn is also specified as an action, message 'EZZ8584I Connection reset by ZERT Policy Enforcement' is written. Otherwise, message 'EZZ8583I Connection logged by ZERT Policy Enforcement' is written. Both messages include a detailed description of the TCP connection characteristics (such as IP addresses, ports, etc) and the cryptographic protection being used for the connection. The message is logged to syslogd using facility LOCAL5. The syslogd priority is determined by the LogLevel parameter.

  Restriction:

  – The syslog daemon must be started. See Syslog daemon in z/OS Communications Server: IP Configuration Reference for more information.

  – TRMD must be started for the affected TCP/IP stack. See Starting the traffic regulation manager daemon (TRMD) from the z/OS shell or Starting the traffic regulation manager daemon (TRMD) as a started task in z/OS Communications Server: IP Configuration Reference for more information.

  Tip: To avoid flooding syslogd, zERT enforcement limits the number of messages that are written to syslogd during a 5-minute interval. See zERT enforcement syslogd message suppression for a description of the message suppression approach. If a complete record of connections matching this ZERT rule is needed, use AuditRecord action.

- Log a message to the console (i.e. TCP/IP job log). This action is configured with the LogConsole parameter and is off by default. If ResetTCPConn is also specified as an action, message EZZ856I CONN RESET BY ZERT POLICY is written. Otherwise, message EZZ8551I CONN LOGGED BY ZERT POLICY is written. Both are multi-line messages that include a detailed description of the TCP connection characteristics (such as IP addresses, ports, etc) and the cryptographic protection being used for the connection.

  Tips:

  – To avoid flooding the TCP/IP joblog, zERT enforcement limits the number of messages that are written to the joblog during a 5-minute interval. See zERT enforcement syslogd message suppression for a description of the message suppression approach. If a complete record of connections matching this ZERT rule is needed, use AuditRecord action.

  – To prevent the TCP/IP job log from growing very large and filling up the spool space, ensure that the TCP/IP job log is being spun-off on a regular basis. See Writing your own master scheduler JCL in z/OS MVS Initialization and Tuning Reference, Determining the source JCL for the started task in z/OS MVS JCL Reference and JESLOG parameter in z/OS MVS JCL Reference for more information.

- Reset the TCP connection. This action is configured with the ResetTCPConn parameter and is off by default.

  Tips:

  – Before enabling the ResetTCPConn action, enable logging or auditing for the rule for a trial period to ensure that no legitimate connections are matching the rule.

– When the ResetTCPConn action is enabled, enable logging to the console (LogConsole) or syslogd (LogSyslogd) as well to provide notification that a TCP connection has been reset.

For a ZERTRule that describes acceptable cryptographic protection, you might want to use the default settings for the actions which silently allow the connection. For a ZERTRule that describes traffic where notification is required, you can enable one or more of the log and audit actions. For a ZERTRule that describes traffic that is unacceptable and should be blocked, the reset TCP connection action and one of the log actions can be enabled.

Additional zERT enforcement policy considerations

Discretionary TLS

Certain applications can turn TLS protection on and off during the lifetime of a connection. This is referred to as "discretionary TLS". Common examples of applications that use discretionary TLS are:

- File Transfer Protocol (FTP): FTP always begins with an FTP 220 response being sent from the FTP server to the client in the clear. After that, a TLS session can be established. The TLS session might be used to protect the user logon and password. The TLS session could then be cleared allowing additional data to be sent in the clear. Or the TLS session could remain in effect throughout the life of the connection.
- Simple Mail Transfer Protocol (SMTP): SMTP can switch TLS on and off by using the AUTH command. An example of a use case is SMTP connecting to different mail servers with different levels of security supported or required.
- Some modes of TN3270E connections negotiate the use of TLS, requiring a small amount of cleartext traffic before TLS is enabled. Once the connection security is enabled, it remains enabled for the life of the connection.
- IBM Sterling Connect:Direct® connections always begin with a small amount of cleartext before connection security (which could be TLS) is enabled. Once the connection security is enabled, it remains enabled for the life of the connection.

In many cases an application that uses discretionary TLS starts in the clear, and then switch TLS on. It might reset TLS security again later and possibly repeat the pattern. Each time a connection changes its security characteristics, zERT is notified and evaluates the connection against your zERT enforcement policy rules.

Your zERT enforcement policy must account for these discretionary TLS applications in both the zERT TLS rule set and the zERT rule set for "No recognized protection" (or None rule set). For example, you cannot ban FTP traffic in the clear because FTP always begins in the clear and can then transition to TLS. You can ensure that if TLS protection is being used for FTP that it is an acceptable TLS version with acceptable cryptographic protection characteristics.

If you have a rule set defined for the None security protocol and if those rules do not generally allow connections with no recognized protection, you need a specific rule with Security Protocol None to allow a discretionary TLS connection to run in the clear without being reset. In most cases, you would also want to avoid any logging or auditing actions for such a rule as well. And, you also need a TLS rule for the same connection to verify that it is using acceptable levels of protection, when it is protected by TLS.

Tip: The Network Configuration Assistant provides checking and guidance for Discretionary TLS applications for your zERT enforcement policy.

zERT enforcement syslogd message suppression

Under certain conditions, zERT enforcement suppresses syslogd messages to avoid flooding the syslogd daemon.

Within a 5-minute interval, a maximum of 10 syslogd messages are written by a single zERT enforcement rule and a maximum of 100 syslogd messages are written across all zERT enforcement rules. An exception is made to ensure that at least one syslogd message is written for every unique zERT enforcement rule that is matched within the 5-minute interval. So, if the 100-message limit has already been reached when a connection maps to a rule with a syslogd logging action and that rule has not yet

been matched during the interval, a message will be written for that connection. Any subsequent matches to that rule during the 5-minute interval will be suppressed.

Logging to syslogd resumes after the 5-minute interval ends. When a connection maps to a rule for which syslogd messages were suppressed in the previous interval, message EZZ8585I is written to syslogd for that rule. It includes the count of suppressed syslogd messages, the name of the zERT enforcement rule, and an indication of whether the rule had a reset TCP connection action or not.

Messages in syslogd can provide an awareness that connections are matching a particular zERT enforcement rule. However, a complete record of matching connections cannot be guaranteed with the syslogd logging action due to built-in message suppression. If a complete record of connections matching a rule is needed, enable the audit action for the rule.

zERT enforcement console message suppression

Under certain conditions, zERT enforcement suppresses console messages to avoid flooding the TCP/IP job log.

Within a 5-minute interval, a maximum of 10 console messages are written by a single zERT enforcement rule and a maximum of 100 console messages are written across all zERT enforcement rules. An exception is made to ensure that at least one console message is written for every unique zERT enforcement rule that is matched within the 5-minute interval. So, if the 100-message limit has already been reached when a connection maps to a rule with a console logging action and that rule has not yet been matched during the interval, a message will be written for that connection. Any subsequent matches to that rule during the 5-minute interval will be suppressed.

Logging to the console resumes after the 5-minute interval ends. When a connection maps to a rule for which console messages were suppressed in the previous interval, message EZZ8564I is written to the TCP/IP job log for that rule. It includes the count of suppressed console messages, the name of the zERT enforcement rule, and an indication of whether the rule had a reset TCP connection action or not.

Messages in the TCP/IP job log can provide an awareness that connections are matching a particular zERT enforcement rule. However, a complete record of matching connections cannot be guaranteed with the console logging action due to built-in message suppression. If a complete record of connections matching a rule is needed, enable the audit action for the rule.

Relationship between zERT enforcement policy and AT-TLS and IP Security policies

At first glance, it might appear that zERT enforcement policy overlaps or even clashes with AT-TLS policy or IPsec policy. While it is true that all these policies focus on cryptographic protection of z/OS IP traffic, the perspective that each policy takes is the differentiator.

AT-TLS and IPsec policies apply cryptographic protection to TCP connections (AT-TLS and IPsec) and other IP traffic (IPsec only) when those connections or traffic flow through the z/OS TCP/IP stack.

zERT policy analyzes the established cryptographic protection of each TCP connection and, if so configured, provides notifications or even defensive actions when specific cryptographic conditions match a zERT enforcement rule. Another key point regarding zERT enforcement is that it applies to all TLS/SSL protection, whether that protection was applied by AT-TLS, JSSE, direct application calls to System SSL, or even the use of third-party TLS libraries like OpenSSL. Similarly, zERT enforcement applies to all SSH protection, whether provided by IBM z/OS OpenSSH or a third party SSH implementation running on z/OS.

Security event reporting: Integrated intrusion detection services

Intrusion is a broad term encompassing many undesirable activities. The objective of an intrusion may be to acquire information that a person is not authorized to have (*information theft*). It may be to cause business harm by rendering a network, system or application unusable (*denial of service*). Or it may be to gain unauthorized use of a system as a stepping stone for further intrusions elsewhere. Most intrusions follow a pattern of information gathering, attempted access and then destructive attacks. Some attacks can be detected and neutralized by the target system. Other attacks cannot be effectively neutralized by the target system. Many of the attacks also make use of spoofed packets which are not easily traceable to their true origin. Many attacks now make use of unwitting accomplices - machines or networks that are

used without authorization to hide the identity of the attacker. For these reasons, detecting information gathering, access attempts and attack accomplice behaviors is a vital part of intrusion detection.

Attacks can be initiated from outside the internal network or from inside the internal network. Particularly vulnerable is an open system such as a public web server or any machine that is placed in service to serve those outside the internal network. A firewall can provide some level of protection against attacks from outside. However, it cannot prevent attacks after the firewall has authorized an external host to communicate with hosts in the internal network, nor can it provide protection in the case where the attack is initiated from inside the network. In addition, end to end encryption limits the types of attacks that can be detected by an intermediate device such as a firewall.

An intrusion detection system can provide detection of some types of attacks. Common intrusion detection system types currently deployed are network sniffers or sensors and vulnerability scanners. Sniffers, placed at strategic points in the network (in front or behind a firewall, in the network, or in front of a host), operate in promiscuous mode, examining traffic real time that passes through on the local network. Sniffers use *pattern matching* to try to match a packet against a known attack which is expressed as an *attack signature*. Sniffers work best against single packet attacks. Limitations are that they cannot deflect the attacking packet, and they cannot evaluate against encrypted data. Scanners do not detect intrusions in real time. They examine a system periodically looking for vulnerabilities or evidence of intrusion. Some scanners evaluate historical data and can identify behavioral anomalies and patterns associated with intrusions.

The z/OS Communications Server provides intrusion detection services (IDS) that enable the detection of attacks and the application of defensive mechanisms on the z/OS server. The focus of IDS is self-protection. IDS can be used alone or in combination with an external network-based intrusion detection system. The IDS is integrated into the z/OS Communications Server stack and can provide the following functions that are unavailable from an external intrusion detection system.

- z/OS CS IDS evaluates data that has been encrypted by IPSec end to end after decryption on the target server system.
- z/OS CS IDS avoids the overhead of per packet examination against a table of signatures for many known attacks. This is accomplished by integrating the attack detection probes into existing error detection logic. This detection is done in real time. IDS policy is examined when an attack is detected to determine the action to be taken.
- z/OS CS IDS detects statistical anomalies real time. Real-time detection is achieved because it is easier for the target system to keep stateful data/internal thresholds and counters.
- z/OS CS IDS implements prevention type of policies that are executed on the system that is the target of the attack. Prevention policies include packet discard and connection limiting.

The IDS is policy driven and the policies are kept in IDS configuration files or LDAP. These policies determine what actions to take for various IDS events. IDS events detected include scans, single packet attacks against the TCP/IP stack, and flooding. Actions include packet discard, connection limiting, and reporting. IDS events can be recorded in syslog files and/or the console. IDS statistics can be recorded in syslog. Packet traces can be taken to document suspicious activities. The TRMDSTAT command provides summary and detailed reporting of IDS events and statistics.

Figure 40 on page 184 shows the z/OS Communications Server IDS architecture.

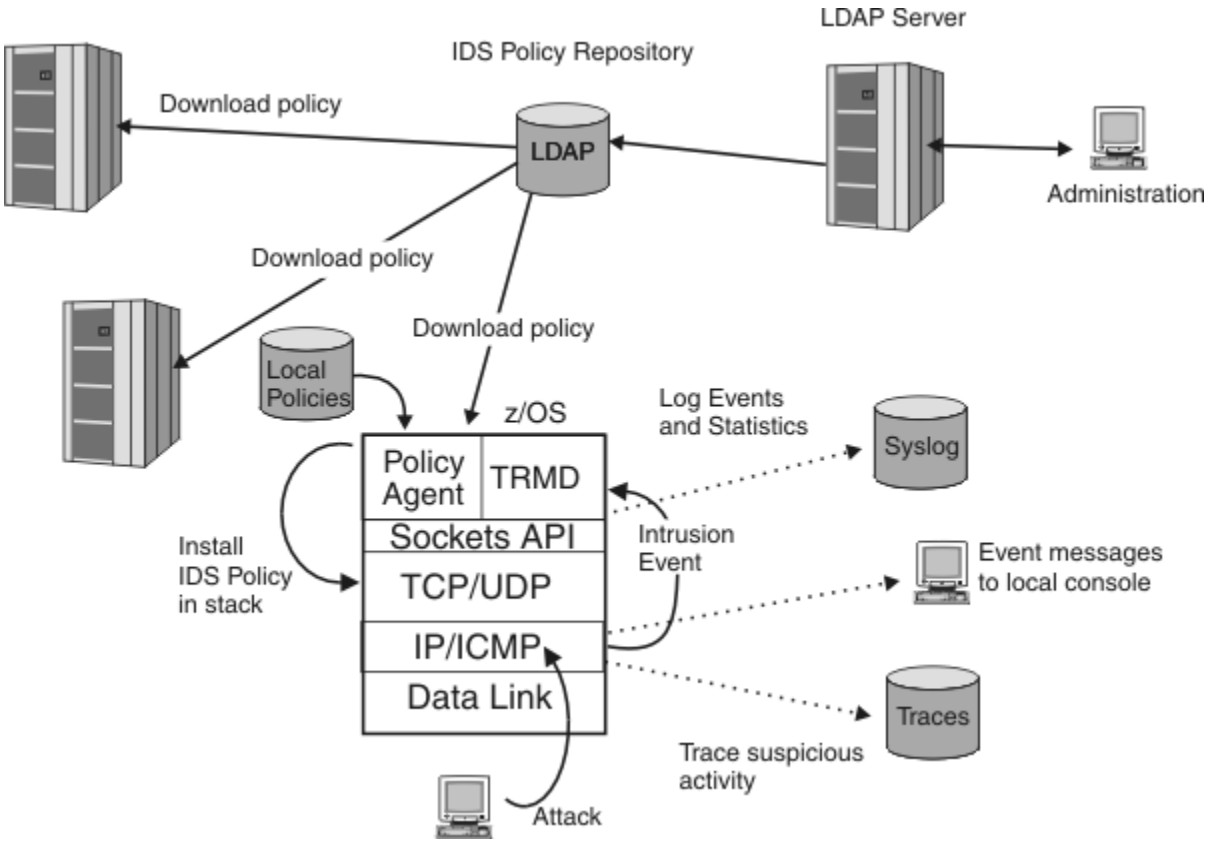


Figure 40. Intrusion detection services overview

For more information on IDS, see Chapter 16, “Intrusion detection services,” on page 851.

Defensive filtering

An external security information and event manager, through analysis and correlation of messages from multiple sources and systems in the network, can take action to block attacks by installing defensive filters in the TCP/IP stack. A defensive filter is a rule to discard packets, separate from IP security filters. Filter processing matches a defensive filter rule to data traffic, based on any combination of IP source or destination address, protocol, source or destination port, or direction of flow. Filter processing checks defensive filters before IP security filters.

The z/OS UNIX **ipsec** command provides the ability to add and manage defensive filters. Defensive filters are typically added automatically as a result of an external security information and event manager's analysis. However, you can also add a defensive filter by manually issuing the **ipsec** command. The Defense Manager daemon (DMD) is an integral part of managing the defensive filters.

Figure 41 on page 185 shows an overview of defensive filtering and the DMD.

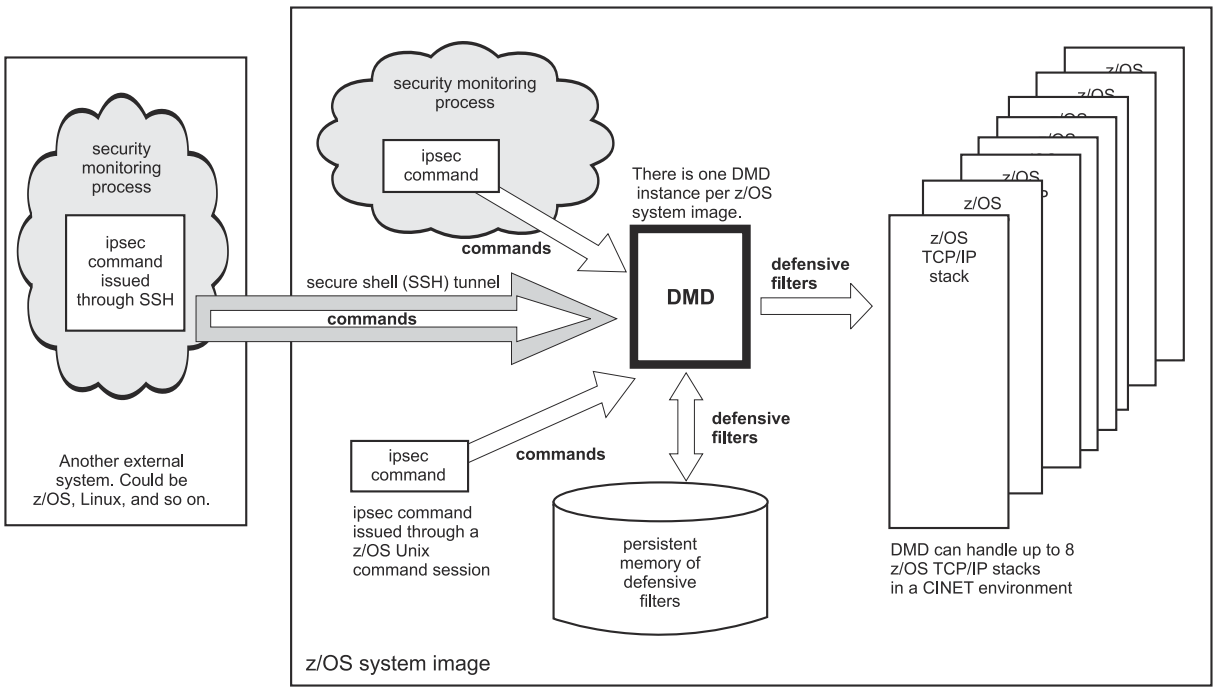


Figure 41. Defensive filtering overview

For more information about defensive filters and the DMD, see Chapter 19, “Defensive filtering,” on page 1107.

Network security services for the IPSec discipline

The network security services (NSS) server provides a set of network security services for the IPSec discipline. These services include a certificate service and a remote management service. The certificate service and remote management service are used by NSS IPSec clients. When an NSS IPSec client uses the certificate service, the NSS server creates and verifies digital signatures on the behalf of the NSS IPSec client. The need to store certificates and keying information at the client, which might be in a less secure zone of the network, is eliminated. When an NSS IPSec client uses the remote management service, the NSS server routes IPSec network management interface (NMI) requests to that NSS IPSec client, which enables the NSS IPSec client to be managed remotely. The NSS IPSec client provides the NSS server with responses to these requests.

You can configure the IKE daemon to act as an NSS IPSec client on behalf of multiple TCP/IP stacks. Each stack appears as a separate NSS IPSec client to the NSS server. Use the `-z` option on the **ipsec** command or use the IPSec NMI to manage NSS IPSec clients that use the remote management service provided by the NSS server. For details about using the **ipsec** command to manage NSS IPSec clients, see z/OS Communications Server: IP System Administrator's Commands. For details about using the IPSec NMI to manage NSS IPSec clients, see z/OS Communications Server: IP Programmer's Guide and Reference.

An NSS IPSec client requires a SAF user ID on the NSS server system. To use the services provided by the NSS server, this user ID must have read access to specific SERVAUTH resource profiles. The following SERVAUTH resource profiles apply to an NSS IPSec client:

- EZB.NSS.*sysname.clientname*.IPSEC.CERT

  This profile authorizes an NSS IPSec client to access the IPSec certificate service of the NSS server.
- EZB.NSS.*sysname.clientname*.IPSEC.NETMGMT

  This profile authorizes an NSS IPSec client to use the IPSec remote management service of the NSS server. The IPSec remote management service of the NSS server enables an NSS IPSec client to be managed by the NSS server.

- EZB.NSSCERT.*sysname*.*mappedlabelname*.HOST

  This profile authorizes an NSS IPSec client to access a personal certificate or a site certificate on the key ring of the NSS server. A profile entry is required for each personal certificate or site certificate associated with an NSS IPSec client. The certificates are used during a phase 1 security association negotiation that uses the digital signature mode of authentication.

- EZB.NSSCERT.*sysname*.*mappedlabelname*.CERTAUTH

  This profile authorizes an NSS IPSec client to access a CERTAUTH certificate on the key ring of the NSS server. A profile entry is required for a CERTAUTH certificate that is connected to the key ring of the NSS server if the following conditions are true:

  – That certificate's label is configured on the CaLabel parameter of a RemoteSecurityEndpoint statement within the IPSec policy definitions for an NSS IPSec client. For details about the RemoteSecurityEndpoint statement, see z/OS Communications Server: IP Configuration Reference.
  – An NSS IPSec client wants to include information about the corresponding certificate authority (CA) to a remote security endpoint. The information is part of the default set of CAs sent when the CaLabel parameter is not specified on the matching RemoteSecurityEndpoint statement.

For additional details about the definition of these profiles, see Chapter 18, “Network security services,” on page 1083.

The following SERVAUTH resource profiles apply when you use the `-z` option of the **ipsec** command or the IPSec NMI to manage NSS IPSec clients:

- EZB.NETMGMT.*sysname*.*clientname*.IPSEC.DISPLAY

  This profile authorizes a user ID to issue the **ipsec** command with the `-z` option to obtain information about an NSS IPSec client, or to make IPSec NMI requests to obtain information about an NSS IPSec client.

- EZB.NETMGMT.*sysname*.*tcpname*.IPSEC.CONTROL

  This profile authorizes a user ID to make IPSec NMI requests to modify the IPSec state of a local stack.

- EZB.NETMGMT.*sysname*.*clientname*.IPSEC.CONTROL

  This profile authorizes a user ID to issue the **ipsec** command with the `-z` option to modify the IPSec state of an NSS IPSec client, or to make IPSec NMI requests to modify the IPSec state of an NSS IPSec client.

- EZB.NETMGMT.*sysname*.*sysname*.NSS.DISPLAY

  This profile authorizes a user ID to issue the **ipsec** command with the `-x` option or to make IPSec NMI requests to obtain information about an NSS server. This profile also authorizes a user ID to issue the **nssctl** command to make NMI requests to obtain information about an NSS server.

For additional details about the definition of the profiles to use the **ipsec** command to manage NSS IPSec clients, see z/OS Communications Server: IP System Administrator's Commands. For additional details about the definition of the profiles to use the IPSec NMI to manage NSS IPSec clients, see z/OS Communications Server: IP Programmer's Guide and Reference.

The following SERVAUTH resource profile applies when the `-w` option of the **ipsec** command or the IPSec NMI is used to display information about an IKE daemon's NSS IPSec clients:

- EZB.NETMGMT.*sysname*.*sysname*.IKED.DISPLAY

  This profile authorizes a user ID to issue the **ipsec** command with the `-w` option or make IPSec NMI requests to obtain information about an IKE daemon's NSS configuration.

For additional details about the definition of this profile to use the **ipsec** command to manage NSS IPSec clients, see z/OS Communications Server: IP System Administrator's Commands. For additional details about the definition of this profile to use the IPSec NMI to manage NSS IPSec clients, see z/OS Communications Server: IP Programmer's Guide and Reference.

Before accessing an IPSec network security service, an NSS IPSec client must present a valid credential. A valid credential consists of the user ID representing the NSS IPSec client and a valid password or

PassTicket. For additional information about using a PassTicket, see z/OS Security Server RACF Security Administrator's Guide.

Certificates and private keys that represent an NSS IPSec client are stored on a single SAF key ring. The NSS server must have access to the certificates on this key ring. The private key associated with a certificate is never sent to the NSS IPSec client. When an NSS IPSec client uses the authentication mode of a digital signature, the NSS server creates a digital signature on the client's behalf. Before the NSS server accesses a personal certificate to create a digital signature on behalf of an NSS IPSec client, the NSS server verifies that the NSS IPSec client is authorized to use that certificate (and its associated private key) by checking the EZB.NSSCERT.*sysname.mappedlabelname*.HOST profile.

Certificates of certificate authorities that are supported by NSS IPSec clients must also be stored on this single key ring. The NSS server provides NSS IPSec clients with information about certificate authorities to which the client has access. The NSS IPSec client advertises certificate authority information to its peer when digital signature mode is used for authentication. The EZB.NSSCERT.*sysname.mappedlabelname*.CERTAUTH profile identifies which certificate authorities an NSS IPSec client can advertise to its IKE peers. For information about this profile name, see “NSS server certificate label naming considerations” on page 1092.

Figure 42 on page 188 shows an example of the IKE daemon acting as an NSS IPSec client for a single TCP/IP stack. The IKE daemon's configuration file identifies which network security services are to be used by a stack, as well as the information about the NSS server from which these services are to be obtained. If the NSS certificate service is enabled for a stack, the IKE daemon uses this service during a phase 1 negotiation when a digital signature mode of authentication is used. If the NSS IPSec remote management service is enabled for a stack, the IKE daemon responds to management requests initiated from the NSS server.

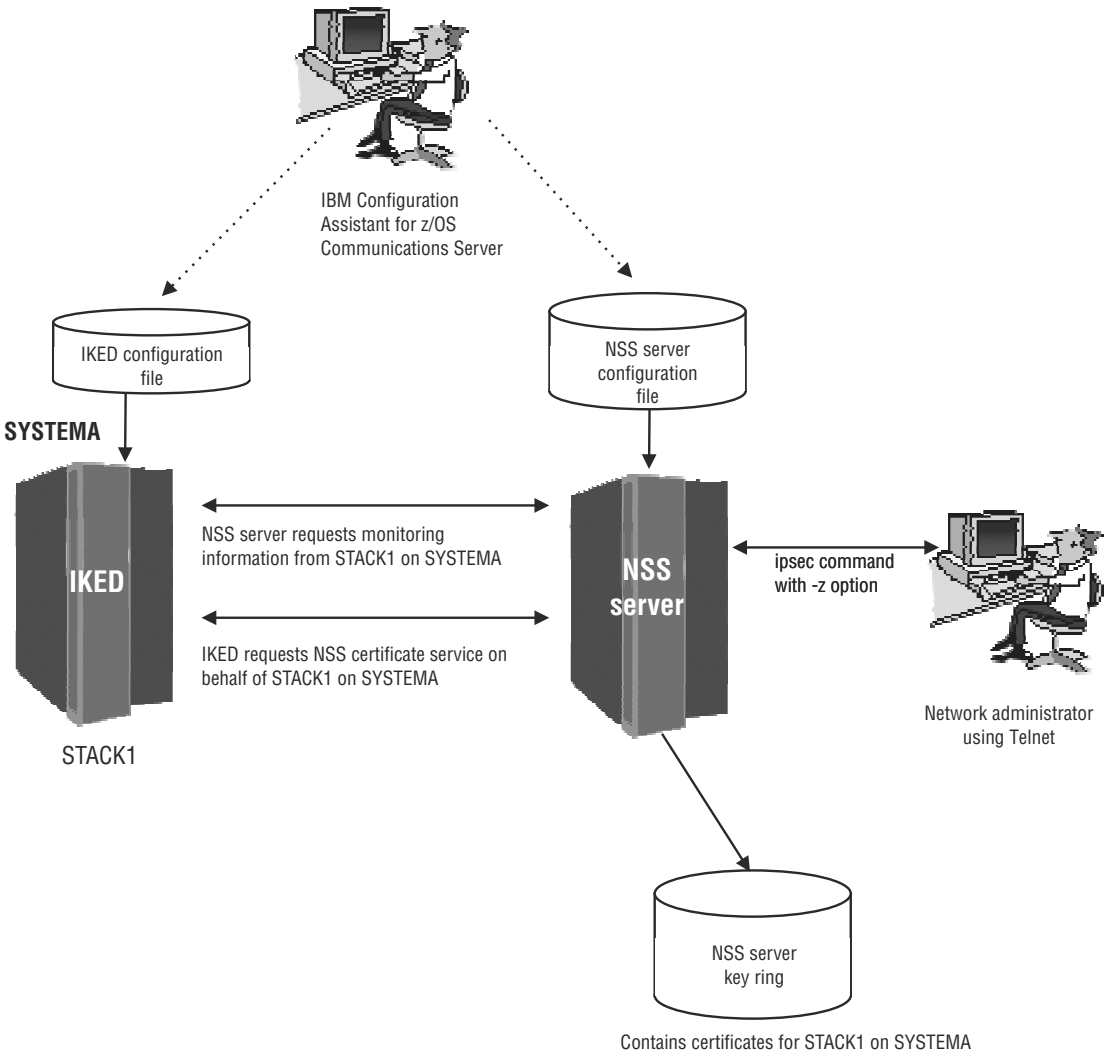


Figure 42. IKE daemon acting as an NSS client for a single TCP/IP stack

The example in Figure 42 on page 188 is a trivial example involving one TCP/IP stack acting as an NSS client. As shown in Figure 43 on page 189, the NSS server can provide services to many NSS clients. The IKE daemon on a z/OS system can act as an NSS client for up to eight TCP/IP stacks (that is, one for each stack running on that system). Multiple IKE daemons can simultaneously access network security services. These IKE daemons do not have to be within the same sysplex as the NSS server.

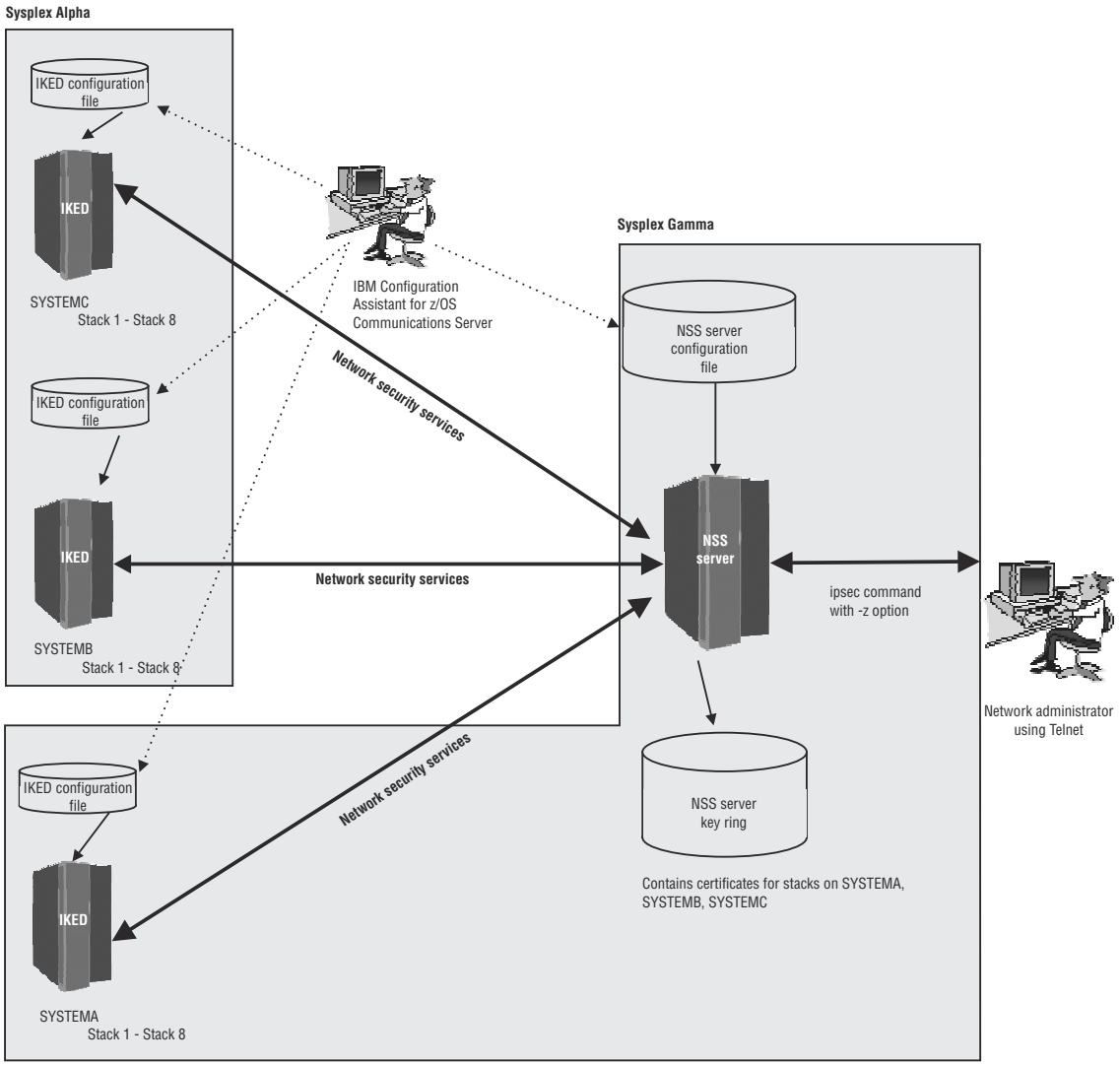


Figure 43. IKE daemon acting as an NSS client for multiple TCP/IP stacks

Network security services for the XMLAppliance discipline

The network security services (NSS) server provides a set of network security services for the XMLAppliance discipline. Services include the SAF access service, the certificate service, and the private key service. NSS XMLAppliance clients can use the network security services in the XMLAppliance discipline. When an NSS XMLAppliance client uses the XMLAppliance SAF access service, the NSS server performs SAF user authentication and access control checks on behalf of the NSS XMLAppliance client. The XMLAppliance certificate service allows an authorized NSS XMLAppliance client to list and retrieve certificates on the configured key ring of the NSS server. The XMLAppliance private key service allows an authorized NSS XMLAppliance client to retrieve private keys that are stored in RACF, generate digital signatures using private keys protected by Integrated Cryptographic Service Facility (ICSF), and perform decryption using ICSF-protected private keys. The NSS server does not support retrieval of ICSF-protected private keys. The NSS server uses its z/OS SAF database to protect unauthorized access to individual certificates and private keys.

Restrictions:

- Because support for digital signature generation and decryption requires the use of ICSF-protected private keys, the applicable Crypto Express2 feature must be defined as a coprocessor, not as an accelerator.
- The XMLAppliance private key service requires that public and private keys be associated with X.509 certificates.

An NSS XMLAppliance client requires a SAF user ID on the NSS server system. To use the XMLAppliance services provided by the NSS server, this user ID must have read access to SERVAUTH resource profiles for each XMLAppliance service. The following SERVAUTH resource profiles apply to an NSS client using XMLAppliance services:

- EZB.NSS.*sysname.clientname*.XMLAPPLIANCE.SAFACCESS

  This profile authorizes an NSS XMLAppliance client to access the XMLAppliance SAF access service of the NSS server.

- EZB.NSS.*sysname.clientname*.XMLAPPLIANCE.CERT

  This profile authorizes an NSS XMLAppliance client to access the XMLAppliance certificate service of the NSS server.

- EZB.NSS.*sysname.clientname*.XMLAPPLIANCE.PRIVKEY

  This profile authorizes an NSS XMLAppliance client to access the XMLAppliance private key service of the NSS server.

- EZB.NSSCERT.*sysname.mappedlabelname*.HOST

  This profile authorizes an NSS XMLAppliance client to access a certificate on the key ring of the NSS server. Use the .HOST or .CERTAUTH profile to authorize an NSS XMLAppliance client to list or retrieve a particular certificate.

- EZB.NSSCERT.*sysname.mappedlabelname*.CERTAUTH

  This profile authorizes an NSS XMLAppliance client to access a certificate on the key ring of the NSS server. Use the .CERTAUTH or .HOST profile to authorize an NSS XMLAppliance client to list or retrieve a particular certificate.

- EZB.NSSCERT.*sysname.mappedlabelname*.PRIVKEY

  This profile authorizes an NSS XMLAppliance client to access the private key associated with a certificate on the key ring of the NSS server. Access to the private key is allowed only if all of the following conditions are true:

  – The private key of the certificate is stored in RACF
  – The private key of the certificate is not stored in the ICSF public key data set (PKDS)
  – The ring usage of the certificate is personal
  – The user ID of the NSS XMLAppliance client has read access to the appropriate certificate label profile (EZB.NSSCERT.*sysname.mappedlabelname*.PRIVKEY).

- EZB.NETMGMT.*sysname.sysname*.NSS.DISPLAY

  This profile authorizes a user ID to issue the **nssctl** command to make NMI requests to obtain information about an NSS server.

Tip: You can specify a wildcard in the profiles to reduce the number of profile entries that you must define.

Before accessing the XMLAppliance services, an NSS XMLAppliance client must present a valid credential. A valid credential consists of the user ID that represents the NSS XMLAppliance client and a valid password or PassTicket. For additional information about using a PassTicket, see z/OS Security Server RACF Security Administrator's Guide.

You control access to certificates and private keys using SAF profiles. The profile name contains a mapped label name that represents the label of the certificate. For information about this profile name, see “NSS server certificate label naming considerations” on page 1092.

Chapter 4. Preparing for IP networking in a multilevel secure environment

Use this information to configure a z/OS Communications Server stack and applications in a multilevel secure environment.

Tip: Many of the tasks, examples, and references in this information assume that you are using the z/OS Security Server (RACF). References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on task performance.

The documentation for many of the applications described in this information references the EZARACF member of sample data set SEZAINST. For examples of the RACF commands used for multilevel security, see the EZARACFM member of sample data set SEZAINST.

Understanding multilevel security concepts

IBM's multilevel security for z/OS has access control implications for your entire system. z/OS Communications Server TCP/IP is only one element of a multilevel secure z/OS system environment. Before planning your multilevel secure IP network, you should be familiar with the concepts and terminology presented in z/OS Planning for Multilevel Security and the Common Criteria.

Multilevel secure networking

The security administrator is responsible for defining the security levels and categories required in a multilevel secure environment. These become part of the set of security labels used to enforce mandatory access control policies when applications access resources on behalf of users. All of the systems enforcing mandatory access control policies in a multilevel secure network must have equivalent definitions of these security labels and the systems in the network to which they apply.

In the networking environment, the information that is being protected is the data being read and written through sockets. Sockets are opened and used by applications running under user IDs. In a z/OS multilevel secure environment:

- Each user ID is permitted to use one or more security labels.
- Every job or login session is associated with a user ID.
- A user ID can use only one security label for each job or login session.
- The security label used is limited by the port of entry (source type and location) of the job or login session.

Applications can have read access to information from many sources that can have various security labels. This information might be commingled in the buffers used to write information to the network. TCP/IP treats all socket data buffers as having the security label of the writing task. All sockets are inherently read/write, so TCP/IP requires communicating partners to have equivalent security labels in a multilevel secure environment.

Nonsecure systems

Most systems do not support mandatory access control processing. If these systems are not physically managed, they normally should not participate in a multilevel secure network. Some installations might need to permit them to participate. In these cases, it is recommended that they be assigned a security label with the lowest security level and a single security category that is not common with any other security label.

© Copyright IBM Corp. 2000, 2022

Managed systems

Systems that do not support mandatory access control processing can participate in a multilevel secure network, if they are physically managed to guarantee that all information on the system has the same single security label and all users of the system are permitted to that security label. These systems are referred to as single-level security or managed systems in this information. This management requires both physical control of the systems and careful management of the network. Managed systems must be prevented from communicating with other managed systems that do not have equivalent security labels.

Systems that support mandatory access control and are configured to implicitly associate the correct security label with each managed system can also communicate with managed systems. The systems that perform mandatory access control are responsible for ensuring that only information from applications with an equivalent security label is sent to a managed system, and that information received from a managed system is given only to applications with an equivalent security label.

Multilevel secure systems

Some systems in the network provide multilevel secure environments. These systems have mechanisms to associate security labels with information accessed through the system and with users logged into the system. The system enforces mandatory access control policies to ensure appropriate separation of information.

Other systems cannot normally associate a single security label with IP addresses owned by a multilevel secure system. The packets being sent from a single IP address on the multilevel secure system might have originated from applications running under different security labels.

Applications on a multilevel secure system can securely communicate with applications on a managed system. Mandatory access control enforcement occurs only on the multilevel secure system. The multilevel secure system is responsible for ensuring that it sends only information from an application with an equivalent security label to any managed system. It also is responsible for ensuring that information received from a managed system is delivered only to an application with an equivalent security label.

When two applications on multilevel secure systems communicate, the security label of the sending application must be communicated to the receiving system so that the receiving system can enforce mandatory access control prior to delivering the information to an application. One mechanism to accomplish this is for the sending system to pass the security label with each packet. The two multilevel secure systems must share a common definition and representation of the security labels that are passed.

z/OS Communications Server TCP/IP stacks on z/OS multilevel secure systems

A z/OS CS TCP/IP stack running in a z/OS multilevel secure environment can optionally be configured as either a restricted stack or an unrestricted stack. A restricted stack is configured with a user ID that is defined with a security label other than SYSMULTI. An unrestricted stack is configured with a user ID that is defined with a security label of SYSMULTI. A single z/OS system can concurrently run up to eight z/OS CS TCP/IP stacks, which can be any mix of restricted and unrestricted stacks.

In either mode of operation, appropriate mandatory access control processing is performed at the transport layer. z/OS Communications Server stacks can be host systems on trusted subnetworks. z/OS Communications Server stacks do not perform mandatory access control processing at the link or network layers, so security labels are not considered in packet forwarding with the exception of sysplex distributor, as described in “Configuring stack sysplex features in a multilevel secure environment” on page 198. Packets that contain security labels are not forwarded by a restricted stack. These packets are discarded by the restricted stack.

Restricted stacks

In this mode of operation, the stack ensures that all sockets are opened by applications running with a security label that is equivalent to the security label of that stack. This guarantees that all information

sent by the stack can be implicitly associated with that stack's security label. The stack also ensures that all information received from the network and delivered to an application is equivalent to the stack's security label. A restricted stack can be viewed by other multilevel secure systems as if it were a managed system.

It is important to note that even though a restricted stack is running under a specific security label, it still receives packets from the network with information covered by different security labels. A restricted stack discards this information rather than deliver it to any local applications. However, this information can appear in storage dumps, logic traces, and packet traces. Diagnostic information should always be protected under the highest security label (SYSHIGH).

Unrestricted stacks

In this mode of operation, the stack allows sockets to be opened by applications with any security label. The stack supports mandatory access control processing that allows its applications to communicate securely with any other managed system or restricted stack.

For applications on unrestricted stacks to communicate securely with each other, Communications Server must be able to determine the security label of the sending application. Unrestricted stacks are permitted to define VIPAs in network security zones with security labels other than SYSMULTI. When one of these is used as either the source or destination IP address of a packet, it implicitly identifies the security label of the information. When both IP addresses in a packet are in security zones with the SYSMULTI security label, the application security label must be explicitly transmitted in the packet. This is known as packet tagging.

Communications Server implements a proprietary form of packet tagging to pass the security label of the sending application to the receiving stack for mandatory access control enforcement against the receiving application. Because of this proprietary format, communications between applications on unrestricted stacks that require packet tagging are supported only when both of those stacks are on the same z/OS system and are communicating over an IUTSAMEHOST link, or are members of the same z/OS sysplex and are communicating over an XCF link.

Stack recognition of a multilevel secure environment

You can activate the SAF SECLABEL class and define security labels on SERVAUTH profiles. This causes the security server to enforce mandatory access control policies for those resources without fully activating a multilevel secure environment. The z/OS Communications Server stack does not perform its extra mandatory access control policy enforcement until you issue the RACF command SETROPTS MLACTIVE. When running with SETROPTS NOMLACTIVE, you should not use unrestricted stacks or define network security zones with a SYSMULTI security label.

When a NetAccess statement is encountered in TCPIP profile processing and MLACTIVE has been set, the stack activates extra mandatory access control policy enforcement in both restricted and unrestricted stacks as follows:

- New sockets are allowed only if a STACKACCESS profile covers this stack.
- Network access is allowed only to IP addresses that are mapped into network security zones covered by NETACCESS profiles.
- Restricted stacks do not normally allow SYSMULTI tasks to have network access to security zones with security labels that are not equivalent to the stack's security label. For more information, see “Exempting certain users of certain programs from full Network Access Control” on page 198.
- Unrestricted stacks transmit packet labels both internally and externally to enable an extra mandatory access control check, between the sending task's security label and the receiving task's security label, when both IP addresses are in security zones with a SYSMULTI security label.
- Distributing stacks consider security labels in choosing target applications.
- TN3270E Telnet servers consider security labels in mapping connections to LU names.
- Internal configuration consistency checks are performed whenever PROFILE.TCPIP or certain SERVAUTH class profile changes are made.

Common INET in a multilevel secure environment

When you start several TCP/IP stacks under OMVS, you are using the Common INET (CINET) PFS. Users and jobs can optionally establish affinity to a single stack, or they can allow CINET to choose a stack. If stack affinity is not set, CINET replicates the socket() command to all stacks attached to it. If a job or user does not have READ access to any of the attached stacks, RACF might generate audit failure messages for those stacks. As long as at least one stack accepts the socket() command, CINET will return success to the application. CINET then routes subsequent commands on that socket to one or more of the stacks that accepted the socket() call. For further information on CINET, see z/OS UNIX System Services Planning.

Network security zones

A network security zone is an administrative name for a collection of systems that require the same access control policy. IP addresses are used to map systems into security zones. This requires that the IP addresses used in your multilevel secure network be predictably associated with a single system or group of systems with the same access control policy. A network security zone can contain a single IP address or any combination of IP addresses and subnetworks. All of the IP addresses in a security zone must have the same security label, though all IP addresses with the same security label do not have to be in the same security zone.

IBM zEnterprise System ensemble

An IBM zEnterprise System (zEnterprise) node participating in an ensemble has access to the following networks:

- Intraensemble data network (IEDN)

  The IEDN is accessed through OSA-Express for zBX (OSX) interfaces and through a HiperSockets channel path that is configured to use the Internal Queued Direct I/O extensions (IQDX) function.
- Intranode management network (INMN)

  The INMN is accessed through OSA-Express for Unified Resource Manager (OSM) interfaces.

The IEDN should be treated the same way as any other data network in a multilevel secure environment. The INMN is handled differently.

The INMN is for only IPv6 and provides a path for some authorized zEnterprise applications, such as those providing platform performance management functions, to communicate with another authorized system performing platform management. For more information about these applications, see *IBM z Systems Ensemble Planning Guide*.

All traffic flowing over an OSM interface is protected by OSM access control rather than network access control. The INMN uses automatically generated IPv6 link-local addresses, so individual systems do not have statically assigned addresses. OSM interfaces provide isolation so that traffic can flow only between a stack connected to an OSM interface and the authorized system performing platform management.

Where your z/OS systems fit in your network

z/OS systems that are not configured with RACF SETROPTS MLACTIVE must be physically managed, as any other managed system.

z/OS systems at V1R5 or later that are configured with RACF SETROPTS MLACTIVE, have appropriate TCP/IP configuration, and have appropriate RACF SERVAUTH class profiles defined, can be placed in trusted subnetworks. Firewalls can allow any managed subnetworks to communicate with these trusted subnetworks. The trusted subnetworks will often be defined as a SYSHIGH security zone and will likely contain several individual IP addresses in other security zones, depending on the mix of restricted and unrestricted stacks within the trusted subnetwork.

Planning stacks on your z/OS systems

Before you begin, determine what z/OS systems need to participate in your multilevel secure network. For each z/OS system, determine what release level it will be running, what sysplex it is a member of, which other multilevel secure systems it needs to communicate with, and what security labels will be used.

The following subtopics include some references to additional information; more information is in z/OS Communications Server: IP Configuration Reference and z/OS Security Server RACF Security Administrator's Guide.

Required configuration in a multilevel secure environment

Some configuration statements that are optional in a discretionary security environment are required in a multilevel secure environment. The default behavior of the stack in a discretionary security environment is to permit most applications when these statements are not defined. The default behavior of the stack in a multilevel secure environment is to fail every application when these statements are not defined. Every stack must have an EZB.STACKACCESS profile in the SAF SERVAUTH class. All referenced IP addresses, except intranode management network (INMN) link-local addresses, must be mapped into security zones by NetAccess statements in the TCPIP profile. A profile that covers the resource EZB.FTP.*sysname*.*ftpdaemonname*.ACCESS.HFS must be defined in the SAF SERVAUTH class for file system access by FTP users. The EZB.SOCKOPT profile must be defined for the following options in the SAF SERVAUTH class:

- IPV6_DSTOPTS
- IPV6_HOPOPTS
- IPV6_NEXTHOP
- IPV6_PKTINFO
- IPV6_RTHDR
- IPV6_RTHDRDSTOPTS
- IPV6_TCLASS
- SO_BROADCAST

In addition, if setting a hop limit greater than the default hop limit, the EZB.SOCKOPT profile must also be defined for the following options in the SAF SERVAUTH class:

- IPV6_HOPLIMIT
- IPV6_MULTICAST_HOPS
- IPV6_UNICAST_HOPS

For more information, see “TCP/IP resource protection” on page 124.

Considerations for IPv6-enabled stacks

IPv6 architecture provides for a plug-and-play environment by automatically generating IP addresses when they are not manually configured. To guarantee predictable IP address association with specific systems, you must disable this capability on z/OS Communications Server stacks through manual configuration.

The stack automatically generates link-local IP addresses using the link-local prefix and the interface ID. To make these addresses predictable, you must manually configure the INTFID parameter on all IPv6 INTERFACE statements. If you are enabling dynamic XCF, you must also manually configure the DYNAMICXCF INTFID parameter on the IPCONFIG6 statement.

Some IPv6 interface types support router autogeneration. When this is enabled, the responsible router informs the stack about prefixes that are supported and the stack generates IP addresses from those prefixes and the interface ID. You must disable this capability on all interfaces that support it, by manually configuring at least one prefix or address using the IPADDR parameter on the INTERFACE statement.

These considerations for IPv6-enabled stacks do not apply to the IBM zEnterprise System (zEnterprise) intranode management network (INMN) that is accessed through OSA-Express for Unified Resource Manager (OSM) interfaces. Traffic over that network is protected by OSM access control rather than network access control.

Deciding whether to use restricted or unrestricted stacks

Many installations will find it easier to run a single unrestricted stack on each z/OS system. The following situations might require you to run one or more restricted stacks on a given system:

- Your installation has administrative requirements that prohibit the use of stacks that allow sockets with different security labels.
- You must allow applications to communicate with applications on a multilevel secure system that is not a z/OS system.
- You must allow applications to communicate with applications on a z/OS multilevel secure system that is not z/OS V1R5 or later.
- Your stacks are not on the same system and are not members of the same sysplex.

Configuring a restricted stack

To configure a restricted stack, take the following actions:

1. Determine the appropriate security label for this restricted stack. Associate a user ID with the stack job, and permit the user ID to the intended stack security label. Make that security label the default security label in the user ID profile.
2. Define a STACKACCESS profile for this stack in the SERVAUTH class with the same security label. This profile might often be UACC(READ).
3. Determine the interface and VIPA addresses needed for this stack.
4. Define one or more security zone names for this stack.

   If you have discretionary access control policies that require different treatment for some of the IP addresses on this stack, they will need to be in different security zones.
5. Define NETACCESS profiles for this stack in the SERVAUTH class with the same security label.

   If your access control policy requires only mandatory access control enforcement, this profile can be generic with respect to the z/OS system name and the TCP stack job name. It might often be UACC(READ).
6. Define a NETACCESS statement that maps this system's external IP addresses into security zone names. This can be placed in a shared data set and included in the PROFILE.TCPIP of other z/OS CS systems in the network.
7. If you define a SOURCEVIPA address for this stack, it must be an IP address in a security zone covered by a NETACCESS profile with the stack's security label.

Configuring an unrestricted stack

To configure an unrestricted stack, take the following actions:

1. Unrestricted stacks must run with the SYSMULTI security label. Associate a user ID with the stack job, and permit the user ID to the SYSMULTI security label. Make SYSMULTI the default security label in the user ID profile.
2. Define a STACKACCESS profile for this stack in the SERVAUTH class with the SYSMULTI security label. This profile might often be UACC(READ).
3. Determine the interface and VIPA addresses needed for this stack.

   If you are running server applications that require multiple instances running under different security labels, you will need at least one VIPA for each security label.
4. Define security zone names for this stack.

If you have discretionary access control policies that require different treatment for some of the IP addresses on this stack, they will need to be in different security zones.

VIPAs defined for use with specific security labels will need to be in separate security zones.

5. Define NETACCESS profiles for this stack in the SERVAUTH class.

   Profiles created for VIPAs used with specific security labels are defined with the appropriate security label. If your access control policy requires only mandatory access control enforcement, these profiles can be generic with respect to the z/OS system name and TCP stack job name. They might often be UACC(READ).

   Profiles for other zones on this stack are defined with the SYSMULTI security label. If there are z/OS systems sharing the RACF database that are not members of the same sysplex, or that are not z/OS V1R5 or later, any generic profile should have UACC(NONE). A fully qualified profile might often be UACC(READ).

6. Define a NETACCESS statement that maps this system's IP addresses into security zone names. This can be placed in a shared data set and included in the PROFILE.TCPIP of other z/OS CS systems in the network.

   Tip: If you have multiple unrestricted stacks and you use the OMPROUTE routing daemon, for more information on local address security zones, see "OMPROUTE" on page 203.

7. If you enable SOURCEVIPA on IPCONFIG and IPCONFIG6 statements, in many circumstances you might be able to avoid explicit packet tagging. The source ip address selection algorithm is enhanced to consider security labels. For IPv4 interfaces defined with the LINK statement, the backward search of the HOME list stops at the first VIPA with a security label that is the same as the security label of the application, or with a security label of SYSMULTI. Place IPv4 source VIPAs in the HOME list preceding the physical interface IP addresses that can use them. VIPAs in SYSMULTI security zones should precede those in other zones. For IPv4 interfaces defined with the INTERFACE statement, use the SOURCEVIPAINTERFACE parameter to select the source VIPA. For IPv6 addresses, only addresses that have security labels the same as that of the application or SYSMULTI are considered. Addresses with equal security labels are preferred over those with a security label of SYSMULTI. Place IPv6 VIPAs on VIRTUAL6 INTERFACE statements. Use the SOURCEVIPAINTERFACE parameter on other INTERFACE statements to associate the set of source VIPAs that can be used with each IPv6 interface.
8. If you define TCPSTACKSOURCEVIPA for this stack, to avoid application failures, it must be in a security zone with a SYSMULTI security label.

Steps for configuring global definitions for all stacks

Define a security zone name for the INADDRANY and LOOPBACK addresses. Define a security label and security zone name for all unknown systems in the multilevel secure network.

Procedure

Perform the following steps to configure global definitions for all stacks:

1. Define a security zone name for the INADDRANY and LOOPBACK addresses.
   a) Define a NETACCESS profile for this zone in the SERVAUTH class for each stack. This profile should be specific with respect to the z/OS system name and TCP stack job name, and should have the same security label as the stack job. You are most likely to make this profile UACC(READ).
   b) Define a NETACCESS statement that maps the INADDRANY and LOOPBACK IP addresses of any system into this security zone name. You can place this statement in a shared data set and include it in the PROFILE.TCPIP file of other z/OS systems in the network.
2. Define one security label that has the lowest security level and one category that is not used in any other security labels.

   This security label can then be used for all unknown systems. Mandatory access control access under this security label will be more restrictive than under SYSLOW.

A task using this security label will have R/O access to resources with SYSLOW, W/O access to resources with SYSHIGH, and R/W access to resources with this security label and SYSMULTI. The task will have no access to resources with any other security labels because they will be disjoint.

Any resources created under this security label will be readable only by tasks running under this security label, SYSHIGH, and SYSMULTI. This significantly reduces the risk from unintended and publicly readable or executable SYSLOW resources.

3. Define a security zone name for all unknown systems in the multilevel secure network.
   - Define a NETACCESS profile for this zone in the SERVAUTH class. This profile can be generic with respect to the z/OS system name and TCP stack job name. If your installation supports communications with unknown systems on all z/OS systems, make this profile UACC(READ). If your installation does not support communications with unknown systems on all z/OS systems, make this profile UACC(NONE).
   - Define a NETACCESS DEFAULT statement that maps all unspecified IP addresses into this security zone name. You can place this statement in a shared data set and include it in the PROFILE.TCPIP file of other z/OS systems in the network.

Exempting certain users of certain programs from full Network Access Control

There are certain network administration programs that, to be fully functional, need to be exempted from some aspects of Network Access Control. For instance, the Ping and Traceroute functions test the network path to a destination system. They often need to traverse routers or firewalls that are at IP addresses mapped into security zones that are not normally mandatory access control accessible from a particular restricted stack. ICMP error messages from these systems will not be delivered to the function, if they are not exempted from the Network Access Control check that limits all traffic to security labels equivalent to the stack security label. Also, routing protocol daemons, such as OMPROUTE, frequently need to exchange routing table information with adjacent nodes that are in security zones that might not be mandatory access control accessible from a particular restricted stack. To operate correctly, these programs must also be exempted from some aspects of Network Access Control.

A SYSMULTI user with UPDATE authority to the EZB.STACKACCESS profile will be exempt from the Network Access Control restriction that all traffic must be with partners that are in security zones with security labels that are equivalent to the stack security label or the security label that is associated with the local IP address. Limit this authority to their usage of the programs that must be exempted, which can be accomplished by first specifying UACC(READ) when you are defining the STACKACCESS profiles, and then granting conditional update access to each by specifying the following command:

```
PERMIT stackaccess_profile_name CLASS(SERVAUTH) ID(*) ACCESS(UPDATE) -
WHEN(PROGRAM(ping,oping,tracerte,otracert,omproute))
```

Note: The WHEN(PROGRAM()) conditional access parameter is not supported on profiles in the SERVAUTH class, except where explicitly stated. PERMITs with WHEN(PROGRAM()) on other profiles might be ignored.

Configuring stack sysplex features in a multilevel secure environment

The following considerations apply to stack sysplex features in a multilevel secure environment:

- If TCPSTACKSOURCEVIPA is configured on a stack, the specified VIPA must be in a NetAccess security zone with a security label that is identical to the stack security label.
- If you use job-specific source IP addressing (see SRCIP in z/OS Communications Server: IP Configuration Reference), the specified IP address must be in a NetAccess security zone with a security label that is permitted on the stack and is equivalent to the specified job. If an interface name is used, at least one of the IP addresses configured on that interface must be in a network security zone with a security label that is either SYSMULTI or equal to the specified job.
- If you use destination-specific source IP addressing (see SRCIP in z/OS Communications Server: IP Configuration Reference), the specified IP address must be in a NetAccess security zone with a security

label that is permitted on the stack and is equivalent to the specified destination. If an interface name is used, at least one of the IP addresses configured on that interface must be in a network security zone with a security label that is either SYSMULTI or that is the same as the specified destination.

- For sysplex distributor, the distributing stack must either be an unrestricted stack or a restricted stack with a security label that is the same as all target stacks. The distributing stack will use the security label of the source security zone and the security labels of the active target applications when selecting a target. The distributing stack will also honor SECLBYSYSTEM when the target application is running under SYSMULTI on an unrestricted stack. In an environment using SECLBYSYSTEM, a distributing stack must be on a system where all security labels are active.
- VIPA takeover must be configured only between stacks with the same security label.
- Distribution of connections that require packet tagging are restricted to flowing over XCF or IUTSAMEHOST links. This restriction applies to the route from the client to the distributor, from the distributor to the target server, and from the target server back to the client.

Defining security labels on other profiles in the SERVAUTH class

A z/OS system with RACF SETROPTS MLACTIVE requires all resource profiles defined in the SERVAUTH class to have a security label. All z/OS Communications Server profiles in the SERVAUTH class have EZA, EZB, or IST as the first qualifier. Resource profiles that require meaningful security labels, such as EZB.STACKACCESS and EZB.NETACCESS, are explicitly identified in this information. The following resource profiles can be defined with the SYSNONE security label:

- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_DSTOPTS
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPLIMIT
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_HOPOPTS
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_NEXTHOP
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_PKTINFO
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_RTHDR
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_RTHDRDSTOPTS
- EZB.SOCKOPT.*sysname*.*tcpname*.IPV6_TCLASS
- EZB.SOCKOPT.*sysname*.*tcpname*.SO_BROADCAST
- EZB.FTP.*sysname*.*ftpdaemonname*.ACCESS.HFS
- EZB.RPCBIND.*sysname*.*rpcbindname*.REGISTRY
- EZB.TRCCTL.*sysname*.*tcpname*.OPEN

Some installations might want to define SO_BROADCAST with the SYSLOW security label to further reduce the exposure of data write_down by restricting datagram broadcast to users running with SYSLOW or SYSMULTI. All other z/OS Communications Server resource profiles can be defined with the SYSNONE security label.

Planning your multilevel secure network

Separate your network into security zones. Each subnetwork of physically managed systems should be defined as a single security zone. Several subnetworks with identical security labels and discretionary access control policy requirements can be assigned the same security zone name. Each trusted subnetwork of self-managed multilevel secure systems likely requires several security zones. The trusted subnetwork can also contain physically managed resources, such as routers and network administrator workstations. The trusted subnetwork security zone is likely to require a SYSHIGH security label. Multilevel secure stacks within the trusted subnetwork must have their interface addresses in security zones with the security label of the stack. VIPAs are usually placed in separate subnetworks dedicated to VIPAs and containing no real interface addresses. Multicast addresses, loopback addresses, and unspecified addresses (IPv4 INADDR_ANY and IPv6 in6addr_any) require security zones as well.

The security administrator takes the following actions:

- Defines security labels in the z/OS security server.
- Creates user IDs in the security server with appropriate security labels.
- Defines common discretionary access control policies.
- Defines the security zone name for each required combination of security label and discretionary access control policy.
- Identifies groups of managed machines that belong in each security zone.
- Provides physical security for each group of machines to ensure only users with appropriate security clearance can use them.
- Configures managed machines so that only the network administrator can set IP addresses.

The network administrator takes the following actions:

- Isolates each group of machines into a subnetwork.
- Configures IP addresses on the machines.
- Configures a firewall to limit each group to communicate only with other subnetworks that have the same security label.
- Assigns network security zone names to subnetworks.
- Defines a NETACCESS statement that maps subnetworks and systems into network security zones. This can be placed in a shared data set that can be included in the PROFILE.TCPIP of all z/OS CS stacks in the network.

Planning for interactive UNIX System Services users in a multilevel secure environment

Create a separate home directory for each security label, set stack affinity by security label, and set host and domain name by security label.

Steps for creating a separate home directory for each security label

Interactive users of z/OS UNIX System Services that are permitted to log on with more than one security label must have a separate home directory for each security label.

Before you begin

This approach is similar to the approach shown in z/OS Planning for Multilevel Security and the Common Criteria.

Procedure

Perform the following steps to create a separate home directory for each security label:

1. For each supported security label:
   a) Log on to an administrative user ID with that security label.
   b) Create a directory with the name of that security label under the /u directory.
   c) For each user permitted to that security label, create a home directory under that security label directory.
2. Create a symbolic link in the /u directory using the special value "$SYSSECR/", perhaps named symsecr as follows:

   ```
   ln  -s  "$SYSSECR/"  /u/symsecr
   ```

   Tip: When issuing this command from the shell, use double quotation marks around the $SYSSECR/ string so that the shell does not attempt variable substitution before passing it to the ln command.
3. Define all users' home directories to be '/u/symsecr/*user*' as follows:

```
ALTUSER user OMVS(HOME('/u/symsecr/user'))
```

Results

This approach is useful in many other situations where a different configuration is required for different security labels.

Steps for setting stack affinity by security label

Installations that start several TCP/IP stacks under UNIX System Services common INET (CINET) often find it useful to set stack affinity for most interactive users. You can set the environment variable _BPXK_SETIBMOPT_TRANSPORT to the name of the stack that users are to use during login profile processing.

Procedure

Perform the following steps to set stack affinity by security label:

1. Create a directory for each security label under the /etc directory.
2. Create a profile script file in each directory that contains:

   ```
   export _BPXK_SETIBMOPT_TRANSPORT=stackname
   ```

3. Create a "$SYSSECR/" symbolic link in the /etc directory, perhaps named seclbl as follows:

   ```
   ln -s "$SYSSECR/" /etc/seclbl
   ```

4. Edit the /etc/profile script file and add the following lines:

   ```
   if  test  -f   /etc/seclbl/profile
   then
   .  /etc/seclbl/profile
   fi
   ```

Host and domain name by security label

Installations that start several TCP/IP stacks under UNIX System Services CINET often find it necessary to define a different host name or domain name for interactive users based on their security label. The first decision that needs to be made is what host name and domain name each TCP/IP stack will have. Two common approaches are to:

- Assign each of the stacks a different host name in the same domain name.
- Assign each of the stacks on the same z/OS image the same host name, but define a different domain name for each security label.

These names are then implemented by creating a separate resolver configuration file for each security label.

Steps for creating a separate resolver configuration file for each security label

You can define a different host name or domain name for interactive users based on their security label. Implement these names by creating a separate resolver configuration file for each security label.

Procedure

Use the same directory structure under the /etc directory as follows:

1. Log in as a file system administrator with each security label.
2. Create a copy of your current resolver configuration file in each security label directory using the following command:

   ```
   cp  /etc/resolv.conf  /etc/seclbl/rsv.cfg
   ```

3. Edit these new files as follows:
   a) Change the TCPIPJOBNAME statement to the appropriate stack job name.
   b) Change the HOSTNAME statement to the appropriate host name.
   c) Change the DOMAINORIGIN or first SEARCH statement to the appropriate domain for this security label.

      Tip: For information about interactions between the DOMAINORIGIN and SEARCH statements, see the TCPIP.DATA configuration statements topic in z/OS Communications Server: IP Configuration Reference.
4. Replace your current resolver configuration file with a symbolic link as follows:

```
ln  -s  "$SYSSECR/rsv.cfg" /etc/resolv.conf
```

Planning for applications in a multilevel secure environment

In planning for applications in a multilevel secure environment, you must first understand the multilevel security programming rules that apply to socket applications.

Trusted network administration server applications
Applications that are part of the infrastructure, such as DNS and routing. The delivered information is not sensitive and needs to be accessible across many security labels.

Trusted multi-level secure server applications
Applications with a login process using port of entry, that do all resource access under the client's login identity, and that maintain separation of information accessed by different client tasks, such as FTP and otelnet.

Process applications that change identity must take the following actions:

- Issue _poe() before identity change (can be done by parent, for example, INETD).
- Change identity in parent process before `fork`, `spawn`, or `exec` to ensure interprocess communication (IPC) resources are properly labeled.
- Access any user resources after the identity change.
- Close unnecessary parent resources before `exec`.

Threaded applications that set identity on a thread must take the following actions:

- Issue _poe() on handling thread before _pthread_security_np().
- Ensure all related processing occurs on threads that have the same identity.
- Change identity on thread before `spawn`, `fork`, or `exec`.
- Access any user resources after identity change on the thread.
- Not access any other thread's resources.
- Close all user resources before removing identity from thread.
- Remove user identity from thread before acquiring new work.

All SYSMULTI applications must ensure that they do not put user data on server resources or other user resources. Pay special attention to debugging, logging, or tracing output. Many servers include user level data in output.

Network administration commands and client applications
Local commands, requiring special controls and privileges, that are used to query, configure, or diagnose the network infrastructure.

IBM zEnterprise System (zEnterprise) platform management applications
Authorized zEnterprise applications that perform platform management functions. For more information about these applications, see *IBM z Systems Ensemble Planning Guide*.

General user commands and client applications
Local commands that access resources (including network resources) under the invoking user's security environment.

Network management interfaces
Application programming interfaces that provide sensitive information.

Unsupported applications
Applications that have not been inspected, or have been inspected and are not trusted in a multilevel secure environment.

Configuring z/OS CS applications in a multilevel secure environment

This topic describes configuration for each of the z/OS CS application categories.

Trusted network administration server applications

The following applications are trusted network administration server applications:

- OMPROUTE
- Resolver
- SNTPD
- TIMED
- TRMD
- z/OS syslog daemon (syslogd)
- z/OS UNIX Policy Agent

They can run under a user ID with the SYSMULTI security label in a multilevel secure environment, provided that they adhere to the following configuration instructions.

OMPROUTE

You should run one instance of OMPROUTE for each stack that is using dynamic route configuration. Each instance of OMPROUTE must run under a user ID with SYSMULTI. OMPROUTE communicates with multicast IP addresses. These addresses must be configured into NetAccess security zones. If OMPROUTE must communicate with adjacent nodes that are not in network security zones with security labels equivalent to the security label of the restricted stack or the security label associated with the local IP address, OMPROUTE must run under a user ID that is SYSMULTI and has update authority to the EZB.STACKACCESS resource profile. A SYSMULTI user with UPDATE authority to the EZB.STACKACCESS resource profile is exempt from the restriction that all traffic must be with partners that are in security zones with security labels that are equivalent to the stack's security label or the security label associated with the local IP address. You should carefully protect your routing configuration files to maintain network security. You should consider using any application level security supported by the routing protocol you use. OMPROUTE must be run with stack affinity for the stack that it is servicing.

OMPROUTE uses multicast UDP datagrams to discover adjacent OSPF routing daemons on common subnetworks. Adjacent OMPROUTE instances then establish TCP connections with each other. When two unrestricted stacks running OMPROUTE are attached to a common subnetwork that is neither XCF nor IUTSAMEHOST, adjacency errors will occur. The multicast UDP datagram is successfully transmitted, but the subsequent TCP connection between the two SYSMULTI interface addresses fails because it requires packet tagging. These adjacency failures can be avoided by preventing OMPROUTE from receiving multicast datagrams from partners with which it cannot communicate.

Steps for avoiding adjacency failures
When two unrestricted stacks running OMPROUTE are attached to a common subnetwork that is neither XCF or IUTSAMEHOST, adjacency errors occur. You can avoid these adjacency failures by preventing OMPROUTE from receiving multicast datagrams from partners with which it cannot communicate.

Procedure

Perform the following steps to prevent OMPROUTE from receiving multicast datagrams from partners with which it cannot communicate.

1. Create a network security zone named URXCF for all interface addresses in XCF or IUTSAMEHOST networks on unrestricted stacks:
   a) Define a generic SERVAUTH NETACCESS profile for this zone with the following RACF command:

   ```
   RDEFINE SERVAUTH EZB.NETACCESS.*.*.URXCF UACC(READ) SECLABEL(SYSMULTI)
   ```

   b) Modify the common NETACCESS profile to define the addresses in this zone:

   ```
   NETACCESS
   192.168.10.0/24 URXCF ; xcf subnet perhaps
   10.254.254.0/24 URXCF ; IUTSAMEHOST subnet perhaps
   ENDNETACCESS
   ```

2. Create a network security zone named UROTHER for all interface addresses in other network types on other unrestricted stacks:
   a) Define a generic SERVAUTH NETACCESS profile for this zone with the following RACF command:

   ```
   RDEFINE SERVAUTH EZB.NETACCESS.*.*.UROTHER UACC(READ) SECLABEL(SYSMULTI)
   ```

   b) Prevent the OMPROUTE running for each unrestricted stack from receiving datagrams from this zone with the following RACF command:

   ```
   PERMIT EZB.NETACCESS.*.*.UROTHER CLASS(SERVAUTH) ID(ompurid) ACCESS(NONE)
   ```

   c) Modify the common NETACCESS profile to define the addresses in this zone:

   ```
   NETACCESS
   10.254.1.0/24 UROTHER ; ethernet subnet perhaps
   ENDNETACCESS
   ```

3. Create a network security zone named URLOCAL for all interface addresses in other network types on each specific unrestricted stack. OMPROUTE is permitted to use this local interface to connect to adjacent OMPROUTE daemons on adjacent restricted stacks.
   a) Define a generic SERVAUTH NETACCESS profile for this zone with the following RACF command:

   ```
   RDEFINE SERVAUTH EZB.NETACCESS.*.*.URLOCAL UACC(READ) SECLABEL(SYSMULTI)
   ```

   b) Modify the local NETACCESS profile for each stack to define the local addresses in this zone:

   ```
   NETACCESS
   10.254.1.17/32 URLOCAL ; local address in ethernet subnet perhaps
   ENDNETACCESS
   ```

Resolver

The resolver task is started by z/OS UNIX and runs under the same identity as the OMVS address space. This identity normally has a security label of SYSMULTI. The resolver processes its configuration, system console commands, and its CTRACE under this identity.

The resolver is configured to use a set of files, data sets, and name servers in a given sequence when asked to resolve a name or IP address. For information on configuring the resolver search sequence, see z/OS Communications Server: IP Configuration Reference.

The name resolution process is performed on the requesting thread of execution under the user identity associated with that thread. Each user must have READ access to the files and data sets configured. This is best accomplished by making these UACC(READ) with SECLABEL(SYSLOW). For the resolver to contact a name server on their behalf, each user must also have appropriate STACKACCESS permission and NETACCESS permission to the security zone of the name server.

z/OS UNIX might be configured in a CINET environment with multiple AF_INET Physical File Systems (TCP/IP stacks). In this environment, users and jobs can optionally have affinity for a single stack or allow CINET to choose a stack for them. When stack affinity is not set, CINET replicates some AF_INET calls to all attached stacks. Other calls are routed by CINET to a single stack based on routing information that CINET has extracted from those stacks. The socket() call is one of the calls that is routed to all connected stacks. This might produce RACF Failure Audit messages to any stacks that the resolver user is not permitted to. These messages can be eliminated by setting stack affinity prior to using resolver functions.

SNTPD

You can run as many instances of SNTPD as appropriate for your installation. Time is not considered sensitive information. You can run any SNTPD server under a user identity with a SYSMULTI security label.

TIMED

You can run as many instances of TIMED as appropriate for your installation. Time is not considered sensitive information. You can run any TIMED server under a user identity with a SYSMULTI security label.

TRMD

You should run one instance of TRMD for each stack that has IDS or IP security functions configured. Each instance of TRMD can run under a user ID with the same security label as the stack it is servicing, or with SYSMULTI. TRMD must be run with stack affinity for the stack that it is servicing.

z/OS syslog daemon (syslogd)

You should run one instance of syslogd per z/OS image under a user ID with the SYSMULTI security label. The AF_UNIX socket (default name /dev/log) created by syslogd must have a security label of SYSMULTI, so that applications running under various security labels can log to it.

Syslogd routes application log messages according to filters configured in /etc/syslog.conf. Log messages from applications running under different security labels can be mixed into an output log file by a filter rule. Each output log file created by syslogd should be configured in a directory with a SYSHIGH security label.

z/OS UNIX Policy Agent

You should run one instance of Policy Agent per z/OS image under a user ID with the SYSMULTI security label. The AF_UNIX socket /tmp/unix.str created by Policy Agent must have a security label of SYSMULTI so that applications running under various security labels can connect to it. The user ID that Policy Agent is running under should have READ access to the EZB.STACKACCESS resource profiles of all stacks on the system.

Trusted multilevel secure server applications

The following applications are trusted multilevel secure server applications:

- TN3270E Telnet server
- z/OS UNIX FTP server
- z/OS UNIX REXEC server
- z/OS UNIX rpcbind server
- z/OS UNIX RSH server
- z/OS UNIX Telnet server

These applications can run under a user ID with the SYSMULTI security label in a multilevel secure environment, provided that they adhere to the following configuration instructions.

TN3270E Telnet server

You can run up to eight instances of the TN3270E Telnet server (Telnet), as appropriate for your installation. You can optionally configure the NACUSERID statement to enable Network Access Control for your Telnet, or use the procedure's user ID.

Telnet maps TCP connections to LU names configured in LU groups or supplied by an LU mapping exit. All of the LU names in a single LU group or that are provided by a single exit must have the same security label. Telnet ensures that the mapped LU name has an equivalent security label to the NetAccess profile for the network security zone containing the client. If the security label for the NetAccess profile is SYSMULTI, Telnet ensures that the mapped LU name also has a security label of SYSMULTI. When performing user login authentication, the SNA application should use this LU name as the port of entry TERMID. The LU name profile in the TERMINAL class must have equivalent security label definitions on both the Telnet system and the SNA application system.

z/OS UNIX FTP server

You can run as many instances of the FTP daemon as appropriate for your installation. You can run any FTPD under a user identity with a SYSMULTI security label. A single FTPD can span a mix of restricted and unrestricted stacks in a CINET environment.

To use NetAccess profiles for IPv4 clients, configure PortOfEntry4 SERVAUTH in the FTP.DATA file of the server. If you do not, you must configure profiles in the TERMINAL class covering all IPv4 addresses in your multilevel secure network, with the same security label defined on your corresponding NetAccess profiles.

z/OS UNIX rpcbind server

You should run one instance of rpcbind per z/OS image under a user ID with UID(0) and the SYSMULTI security label. You must define the resource profile EZB.RPCBIND.*sysname*.*rpcbindname*.REGISTRY in the SERVAUTH class. Enable applications to register and unregister with rpcbind by granting at least READ access to the resource profile for the user IDs under which the applications run.

Tip: The registration and deregistration procedures PMAPPROC_SET, PMAPPROC_UNSET, RPCBPROC_SET, and RPCBPROC_UNSET are defined in RFC 1833 *Binding Protocols for ONC RPC Version 2*.

Requirements:

- The rpcinfo utility can list and delete rpcbind registrations. You cannot delete rpcbind registrations using rpcinfo unless the user ID has at least READ access to the resource profile.
- Some RPC library calls register and deregister applications with rpcbind or portmapper. These calls include registerrpc(), svc_register(), pmap_set(), and pmap_unset(). An application that uses these library calls must run under a user ID that has been granted at least READ access to the resource profile.

Client programs can request that rpcbind forward RPCs to registered server programs. For each of these target assistance requests, the rpcbind server forks a new process that runs under the security label of the client's network security zone. The user ID under which rpcbind runs must have at least READ access to each of these security labels that you want to support.

Tip: The RPC library routine pmap_rmtcall() issues target assistance requests on behalf of the caller.

z/OS UNIX INET daemon

The z/OS UNIX INET daemon (inetd) can be configured to listen for requests for many different services. You can run inetd under a user identity with a SYSMULTI security label. Each connection to one of its ports causes it to issue the _poe() service and fork a new instance of the server configured to handle that service. The new connection is passed to the server. If a user ID for the forked server is configured in the inetd configuration file, that user ID must be authorized to the SERVAUTH NETACCESS profile for the

network security zone containing the client IP address. If the forked server performs additional user login authentication, that user ID must also be permitted to the NetAccess profile.

The following z/OS CS servers are designed to be forked by inetd. They each perform additional user authentication.

- z/OS UNIX REXEC server
- z/OS UNIX RSH server
- z/OS UNIX Telnet server

Trusted single-level secure server applications

TSO REXEC and RSH servers are trusted single-level secure server applications.

TSO REXEC and RSH servers

The MVRSHD server listens for both rexec and rsh client connections. Requests are submitted as TSO batch jobs and spool output is returned to the client. MVRSHD does perform client authentication and does isolate user data for each connection. However, it does not request port of entry processing during client authentication, and job submission and spool access are performed under the identity of the MVRSHD job rather than the client identity.

RSH client requests are run under the identity and security label of the MVRSHD job. REXEC client requests are run under the user ID specified on the request.

MVRSHD supports an optional startup parameter to control whether or not it inserts the SECLABEL parameter on the generated job card. When SECLABEL=Y is specified at startup, MVRSHD passes its own security label in the SECLABEL job parameter. The user ID specified on the request must be permitted to the MVRSHD server's security label. When the parameter is not specified, or is specified as SECLABEL=N, the default security label of the user ID specified on the request must be identical to the MVRSHD server's security label.

You must run a separate instance of MVRSHD for each security label you need to support. Run each one under a different job name assigned to a user ID with the appropriate security label. You can run multiple MVRSHD servers on the same unrestricted stack. Define a VIPA in a network security zone with the appropriate security label for each server on that stack. Use the PORT reservation statement in the TCPIP profile to override the bind address of each job to the appropriate VIPA.

Network administration client applications

The following applications are network administration client applications:

- nsupdate
- Netstat
- pasearch
- Ping
- Traceroute
- trmdstat

For security implications on the use of these commands, see the following information on configuring these applications for use in your multilevel secure environment.

nsupdate

The nsupdate command is used to update the name server database. Because the name server is often running with a SYSMULTI security label and listening on IP addresses in SYSMULTI network security zones, mandatory access control policy does not limit a general user's ability to use the nsupdate command to contact the server. File system access control lists and security labels should be used to limit execution of this command to intended network administrators. Application level security options should also be considered.

Netstat

Use the Netstat command to display stack settings, and information about open ports and established connections. You might want to restrict the availability of some Netstat options to network administrators. For information on restricting Netstat options, see "Netstat access control" on page 145.

pasearch

The pasearch command is provided to interrogate the stack policies currently in place on a system. It requires that the user have superuser privileges. Only network and security administrators that need to know policy settings should be permitted to issue this command. To limit access, set the file mode access permission bits using the CHMOD command.

Ping

The Ping command is used to verify the network path to a given destination. It sends ICMP Echo Request datagrams to the destination and listens for ICMP Echo Reply responses. Normal Network Access Control limits a user's ability to send and receive these datagrams. On a restricted stack, all users are limited to sending and receiving datagrams to destinations in network security zones with security labels equivalent to the stack. On an unrestricted stack, users are limited to destinations equivalent to their own security label. Users with a SYSMULTI security label on an unrestricted stack are limited to those security zones with which they are authorized to communicate.

You can permit SYSMULTI users to use the Ping command for IP addresses in security zones with security labels that are not equivalent to that stack by using the RACF PERMIT command to give them UPDATE access to the STACKACCESS profile for that stack. You can configure this permission so that it is in effect only for specific programs invoked by the users. For example, to permit users to invoke the Ping command, specify the WHEN(PROGRAM(OPING,PING)) parameter on the PERMIT command.

Traceroute

The Traceroute function is used to verify the network path to a given destination and identify each intermediate system. It sends UDP datagrams to the destination with increasing hop count values, and listens for ICMP TIME EXCEEDED INTRANSIT and PORT UNREACHABLE responses. Normal Network Access Control limits a user's ability to send and receive these datagrams. On a restricted stack, all users are limited to sending datagrams to destinations in network security zones with security labels equivalent to the stack, and receiving datagrams from intermediate systems in equivalent security zones. On an unrestricted stack, users are limited to destinations equivalent to their own security label. Users with a SYSMULTI security label on an unrestricted stack are limited to those security zones with which they are authorized to communicate.

You can permit SYSMULTI users to trace the route to addresses in security zones that are not equivalent to that stack by PERMITing them with UPDATE access to the STACKACCESS profile for that stack. This PERMIT can be limited to apply only when using certain programs, such as Traceroute, by using the WHEN(PROGRAM(tracerte,otracert)) clause on the PERMIT.

trmdstat

The trmdstat command is provided to process syslogd output files and generate IDS reports. The network or security administrator using this command needs to be authorized to read the syslogd output files. This often requires a user security label of SYSHIGH or SYSMULTI.

IBM zEnterprise System platform management applications

Some IBM zEnterprise System (zEnterprise) authorized applications communicate over the intranode management network (INMN), such as those providing platform performance management functions. For more information about these applications, see *zEnterprise System Ensemble Planning and Configuring Guide*.

These applications are not supported in a multilevel secure environment when run under a user ID with a security label of SYSMULTI. They are supported when run under a user ID with a specific security label.

Some of these applications listen using specific multicast IP addresses and ports to determine the remote IPv6 link-local IP address and port number to use for communication with another authorized system performing platform management.

For the zEnterprise authorized platform management application to connect to the INMN, at least one TCP/IP stack on each z/OS image must be connected to an OSA-Express for Unified Resource Manager (OSM) interface. That stack can be an unrestricted stack, or it can be a restricted stack running with a security label that is equivalent to the security label used for the zEnterprise system management agent application.

You must configure TCP/IP and the security server as follows:

- Permit the application to the STACKACCESS resource profile of the stack with one or more OSM interfaces.
- Define an OSMACCESS resource profile on the stack with the OSM interface, using the same security label as that of the application.
- Permit the application to the OSMACCESS resource profile.
- Define a NETACCESS resource profile for the multicast IPv6 address that is used by the application.
- Permit the application to the NETACCESS profile for that multicast address.

General user client applications

The following applications are general user client applications:

- dig
- dnsdomainname, domainname
- host
- hostname
- MISCSERV (echo, discard, character generator)
- NSLookup
- TSO REXEC client
- TSO RSH client
- TSO Telnet client
- z/OS UNIX FTP client
- z/OS UNIX REXEC client
- z/OS UNIX RSH client

These commands are supported for use by general users in a multilevel secure environment. Mandatory access control policy enforcement limits the stacks that can be used, the servers that can be contacted, and the local resources that can be used. No extra, application specific, multilevel secure environment configuration is required. For discretionary security environment considerations, see the information for a particular application and see z/OS Communications Server: IP Configuration Reference.

Network management interfaces

This topic describes the multilevel secure networking considerations for those network management interfaces (NMIs) that provide sensitive information.

Real-time application-controlled TCP/IP trace NMI

This NMI can be used to obtain packet and data trace data. For more information about this NMI, see the topic Real-time application-controlled TCP/IP trace NMI (EZBRCIFR) in z/OS Communications Server: IP Programmer's Guide and Reference.

In a multilevel secure environment, define the security labels for the NMI RACF resource profiles and for the user IDs of the NMI applications as follows:

RCCOpen request resource
Define the EZB.TRCCTL.*sysname*.*tcpname*.OPEN resource profile with a security label of SYSNONE.

RCCSetFilters request packet and data trace resources
The packet and data trace data provided by the NMI are considered sensitive information that must be secured. The following resource profiles are associated with this data:

- EZB.TRCCTL.*sysname*.*tcpname*.PKTTRACE
- EZB.TRCCTL.*sysname*.*tcpname*.DATTRACE
- EZB.TRCSEC.*sysname*.*tcpname*.IPSEC
- EZB.TRCSEC.*sysname*.*tcpname*.ATTLS

Set the security label associated with these resource profiles to be the same as the security label of the user ID associated with the TCP/IP stack.

User ID of applications
For those applications that request packet or data trace data, the security label associated with the user IDs for the applications can be SYSHIGH (or installation equivalent) if the security label associated with the TCP/IP stack is SYSMULTI or SYSHIGH. However, if the security label associated with the TCP/IP stack is not SYSMULTI or SYSHIGH, the security label associated with the user IDs must be SYSMULTI.

Unsupported applications

The following applications are not supported in a multilevel secure environment:

- HOMETEST command
- LPD
- LPQ command
- LPR command
- LPRM command
- LPRSET command
- NPF
- Portmapper
- SNMP NetView client
- TELNET client command
- TESTSITE command
- TNF
- VMCF
- z/OS UNIX Network SLAPM2 subagent
- z/OS UNIX OMPROUTE SNMP subagent
- z/OS UNIX popper
- z/OS UNIX RSVP agent
- z/OS UNIX SNMP client command
- z/OS UNIX SNMP server and agent
- z/OS UNIX Trap Forwarder Daemon

Any other commands or applications not explicitly mentioned in this information have not been inspected and are not supported in a multilevel secure environment.

Changing your multilevel secure networking environment

Changes to certain parts of your multilevel secure configuration should be well planned. Changes to the NETACCESS statement in PROFILE.TCPIP, the security label associated with the TCPIP started task, the security label associated with the EZB.STACKACCESS or EZB.NETACCESS profiles in the SERVAUTH class, or the definition of profiles in the SECLABEL class can result in an IP address being associated with a different security label. These changes imply that corresponding changes have been implemented that affect the security management of the affected systems. For systems that support mandatory access control policy enforcement, those policies must be updated at the same time. For systems that do not support mandatory access control policy enforcement, the physical user access procedures, system data content, firewall configurations, and router configurations must be updated at the same time.

The safest method of controlling mandatory access control policy changes is to use the RACF options MLSTABLE and MLQUIET. When the RACF options are set to MLACTIVE, MLSTABLE, and NOMLQUIET, TCP/IP does not permit an existing NETACCESS configuration to be changed with a VARY TCPIP,,OBEYFILE command. When the RACF option MLQUIET is set, RACF requires the TCP/IP job user ID to be RACF SPECIAL to open data sets referenced by VARY TCPIP,,OBEYFILE commands. For more information on setting and using these options, see z/OS Security Server RACF Security Administrator's Guide.

In an MLSTABLE environment, all user and application access to data should be halted before entering the MLQUIET environment. All network access can be halted by stopping all TCP/IP stacks. If security administrators must access the system through the TN3270E Telnet server (Telnet), consider running a restricted stack with only Telnet for the security administrators. Permit the Telnet NACUSERID or the procedure's user ID to the EZB.STACKACCESS profile for this stack. Stop all other TCP/IP stacks. After the local policy changes and all coordinated changes on other systems are complete, set NOMLQUIET and restart your production TCP/IP stacks and network servers.

Every stack running on a system with the RACF option MLACTIVE does an internal consistency check on several PROFILE.TCPIP statements and their associated SERVAUTH profiles. This consistency checking occurs at the end of initial profile processing, after the VARY TCPIP,,OBEYFILE command modifies the profile, and whenever RACLIST is issued for the SERVAUTH or SECLABEL classes. Some applications, such as OMPROUTE, also cause the consistency checking to occur because they internally issue an equivalent of the VARY TCPIP,,OBEYFILE command. TCP/IP writes a message to the job log for each inconsistency it finds that could compromise the security of information flowing through the stack. If inconsistencies are found, a final message (EZD1217I) summarizing the number of problems found is written to the system console.

By default, the stack will continue running when inconsistencies are found. It is recommended that you override this default by specifying GLOBALCONFIG MLSCHKTERMINATE in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command, before starting production workloads. Before making security related configuration changes, it is recommended that you first stop all production workloads. You can then specify GLOBALCONFIG NOMLSCHKTERMINATE in PROFILE.TCPIP or in the data set referenced by a VARY TCPIP,,OBEYFILE command. This parameter can be changed only from MLSCHKTERMINATE to NOMLSCHKTERMINATE when the RACF option NOMLSTABLE is set or when both MLSTABLE and MLQUIET are set.

The stack performs the following consistency checks:

- The stack does not currently have a NETACCESS statement configured.
- The TCP/IP job security label must be currently active on this system.
- A STACKACCESS profile for this stack must exist.
- The STACKACCESS profile must have the same security label as the TCP/IP job.
- The NETACCESS statement must have both INBOUND and OUTBOUND turned on.
- A NETACCESS profile must exist for each defined NETACCESS security zone.
- If NETACCESS DEFAULTHOME is configured, it must have a zone with the same security label as the TCP/IP job.

- If NETACCESS TcpStackSourceVIPA is configured, it must be in a zone with the same security label as the TCP/IP job.
- The special address INADDR_ANY (0.0.0.0) must be in a NETACCESS security zone.
- The special address INADDR_ANY (0.0.0.0) must be in a zone with the same security label as the TCP/IP job.
- On an IPv6-enabled stack, the special IPv6 unspecified address [in6addr_any (::)] must be in a NETACCESS security zone.
- On an IPv6-enabled stack, the special IPv6 unspecified address [in6addr_any (::)] must be in a zone with the same security label as the TCP/IP job.
- On an IPv6-enabled stack, every INTERFACE statement must have a manually configured INTFID. (OSA-Express for Unified Resource Manager (OSM) interfaces are exempt from this check.)
- On an IPv6-enabled stack, every INTERFACE that supports autoconfiguration must have at least one manually configured IPADDR. (OSM interfaces are exempt from this check.)
- On an IPv6-enabled stack with dynamic XCF, DYNAMICXCF INTFID must be configured on the IPCONFIG6 statement.
- Every home address on the stack must be in a NETACCESS security zone. (OSM interfaces are exempt from this check.)
- Every home address that is not a VIPA must be in a zone with the same security label as the TCP/IP job. (OSM interfaces are exempt from this check.)
- Every home address that is a VIPA must be in a zone with an equivalent security label as the TCP/IP job.
- On a restricted stack, VIPAs must not be in a security zone with the SYSMULTI security label.

There are several consistency checks that are not performed by the stack. These remain the responsibility of the system security administrator:

- The NetAccess zone definitions must agree across stacks. It is recommended that common definitions from a shared data set be included in all stack profiles.
- Subnetworks or networks are entirely contained in a single NetAccess security zone. NetAccess checks for authorization to a specific destination address, but broadcast packets are delivered to all addresses within a subnetwork or network. Addresses owned by multilevel secure stacks are the only addresses that can be safely configured into a network security zone different than their subnetwork or network.
- RACF definitions must be consistent across RACF data sets. It is recommended that installations use RACF facilities to manage this consistency. SECDATA and SECLABEL class profile consistency is critical to preserve the meaning of SECLABEL profile names. It is suggested that installations define SERVAUTH profiles for all zones that are generic across all systems and stacks, except those containing the loopback address and the unspecified address (IPv4 INADDR_ANY and IPv6 in6addr_any).

Continue making changes to either PROFILE.TCPIP statements or RACF profiles until no consistency errors are reported. Then, specify GLOBALCONFIG MLSCHKTERMINATE in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command. At that point, it is safe to set NOMLQUIET and restart production workloads.

Chapter 5. TCP/IP Customization

Before you begin customizing, it is assumed that you know what configuration data sets are used by the TCP/IP address space, their search order, and considerations for what type of TCP/IP stack you will be running in your environment (for example, Enterprise Extender (EE) and multiple stacks). See Chapter 2, "IP configuration overview," on page 11 for this information.

After reading this information, you will know how to configure and start syslogd and the TCP/IP stack. You should understand the relationships of TCP/IP configuration files as they apply to the TCP/IP address space. The four main configuration files that you will be working with are:

- TCPIP.DATA
- PROFILE.TCPIP
- HOSTS.LOCAL
- ETC.IPNODES

You should be able to use the following commands to verify customization:

TSO PING and z/OS UNIX ping
Sends IP datagrams to a specified destination host, requesting a reply, and measures the round trip time. This helps you to verify the interfaces defined to the TCP/IP address space.

TSO NETSTAT, z/OS UNIX netstat, and DISPLAY TCPIP,,NETSTAT
Queries TCP/IP about the network status of the local host or the contents of the resolver cache. With Netstat, you can verify most TCP/IP customization values that can be set from PROFILE.TCPIP. You can also display detailed information about the contents of the system-wide resolver cache, or statistical information such as the number of cache queries.

TSO HOMETEST
Verifies your host name and address configuration.

TSO TRACERTE and z/OS UNIX traceroute
Displays the route that a packet takes to reach a requested destination.

Configuring the syslog daemon

The syslog daemon (syslogd) processing is controlled by a configuration file called /etc/syslog.conf, in which you define logging rules and output destinations for error messages, authorization violation messages, and trace data. Logging rules are defined using a *facility* name and a *priority* code. For locally generated messages, the user ID and job name of the program that generated the message can also be specified in the rule. For messages arriving over the network, the rule can include the IP address or host name of the sender. The *facility* name and *priority* code are passed on the logging request from an application when it wants to log a message. The user ID and job name are provided by the system. See z/OS Communications Server: IP Configuration Reference for more information about logging rules.

You can specify statements and rules in the configuration file using a variety of EBCDIC code pages. Use the SYSLOGD_CODEPAGE environment variable to specify the code page that you want to use. The default code page is IBM-1047.

As shown in the following sample /etc/syslog.conf file, comments can be added to the configuration file by placing the number (#) character in column one of the comment line. Everything following the number (#) character is treated as a comment. This sample is available in /usr/lpp/tcpip/samples/syslog.conf.

```
#  Licensed Materials - Property of IBM
#  5650-ZOS
 #  Copyright IBM Corp. 2010, 2017 #  Status = CSV2R3
#
# /etc/syslog.conf - control output of syslogd
#
```

© Copyright IBM Corp. 2000, 2022

```
# The # sign begins a comment which extends to the end of the line.
#
# Blank lines are ignored.
#
# See IP Configuration Reference for detailed information about
# the syntax.  These comments are meant to provide only a general
# overview.
#
# There are two types of configuration information:
#
# 1) Global configuration values that control the behavior of syslogd.
# 2) Rules that specify types of messages which syslogd will
#    store, and where syslogd will store them.
#
# Global configuration statements:
# --------------------------------
#
# The following global statements control the syslogd automatic archive
# function.  All statements except BeginArchiveParms should only be
# specified once.  If you specify them multiple times the last
# instance is used.  The BeginArchiveParms statement can be repeated
# multiple times, and each instance pertains to the rules that follow
# it, until another instance is specified.  Each instance completely
# replaces the previous instance.
#
# The automatic archive function archives UNIX file destinations to
# MVS sequential or generation data group (GDG) data sets.  The
# particular UNIX files that are to be archived must include the -N

# parameter.  The -N parameter specifies that the file should be
# automatically archived and then re-initialized when an archive event
# occurs.  Alternatively, you can specify the -X parameter.  The -X
# parameter specifies that the file should only be re-initialized but
# not archived when an archive event occurs.  Use the -X parameter only
# if you do not want to save the contents of the log file.  If you
# don't specify the -N or -X parameter, then the file does not
# participate in automatic archival.

#
# Archival occurs for the following events:
#
# - At the time of day configured on the ArchiveTimeOfDay statement.
# - When one or more UNIX file systems reach the percentage full
#   configured on the ArchiveThreshold statement.
# - When the MODIFY procname,ARCHIVE command is issued.
#
# ArchiveTimeOfDay
#   Specifies the local time of day in hours and minutes using a 24
#   hour clock.  Syslogd archives all eligible files at the specified
#   time of day.  There is no default - if you don't specify this
#   statement, then syslogd does not perform time of day archival.
#
# ArchiveThreshold
#   Specifies the percentage of file system full that triggers an
#   archive.  This value applies to all UNIX file systems represented
#   by the set of UNIX files in all rules.  When any file system
#   reaches the specified percentage full, files in that file system
#   are archived until the percentage full reaches half of the
#   configured value.  Files are archived starting with the largest
#   and working toward the smallest.  For example, if you configure
#   80% full, files are archived until the file system is only 40%
#   full.  The default is 70%.
#
# ArchiveCheckInterval
#   Specifies the value in minutes for checking the percentage full
#   for UNIX file systems.  The default is 10 minutes.
#
# BeginArchiveParms
#   Specifies archive data set details.  The following parameters
#   can be specified.  DSNPrefix is required but all other parameters
#   are optional (although they might need to be specified for your
#   installation in order for archival to succeed).
#
#   DSNPrefix
#     Specifies the data set name prefix for the archive data set.
#     The complete data set name is formed by concatenating this
#     prefix value with the unique qualifier specified on the -N
#     parameter on a particular rule, and with a unique suffix value.
#     See IP Configuration Reference for complete details on the data
#     set name.
#
#   Unit
```

```
#     Specifies the unit for the archive data set.
#
#   Volume
#     Specifies the volume for the archive data set.
#
#   MgmtClas
#     Specifies the management class for the archive data set.
#
#   StorClas
#     Specifies the storage class for the archive data set.
#
#   RetPd
#     Specifies the retention period in days for a sequential archive
#     data set.  Valid values are 0 - 9999.
#
# The following are example statements that illustrate how to
# configure automatic archival.  See the section on syslogd rules for
# details about specifying rules.
#
#   ArchiveTimeOfDay         00:01
#   ArchiveThreshold         70
#   ArchiveCheckInterval     10
#
#   BeginArchiveParms
#     DSNPrefix              USER1.ARCHIVE
#     Volume                 VOL001
#     RetPd                  30
#   EndArchiveParms
#
#   *.SYSLOGD.daemon.notice  /var/logs/syslogd/daemon.notice -N daemon.notice
#
# NOTE: The archive data set name for the above example will be:
#
#   USER1.ARCHIVE.DAEMON.NOTICE.Dyymmdd.Thhmmss
#
# Syslogd rules:
# --------------
#
# Four criteria can be used to select locally generated
# messages for processing:
#
# 1) user ID associated with application generating the message
#
#    * can be specified for the user ID if the user ID is not
#    important.
#
# 2) job name of application generating the message
#
#    * can be specified for the job name if the job name is not
#    important.
#
# 3) facility of the message, as specified by the application
#
#    This is user, mail, news, uucp, daemon, auth, cron, lpr, or
#    local0-local7.  Consult the documentation for the application
#    to determine which facility the application specifies.
#
#    A special facility, mark, specifies that syslogd should log
#    mark messages on a regular basis.  These can be used to verify
#    that syslogd was operational during a specific time interval.
#
# 4) priority of the message, as specified by the application
#
#    This is emerg, panic, alert, crit, err, error, warn, warning,
#    notice, info, or debug. A filter rule condition using a specified
#    priority will match messages with that priority or higher; higher
#    meaning more severe.
#
#    A special priority, none, specifies that messages with the
#    specified user ID, job name, or facility should not be
#    selected.
#
# These criteria are specified together as
#
#    userid.jobname.facility.priority
#
# or, if user ID and job name are both *, as
#
#    facility.priority
#
# This can be combined in a series as
#
```

```
#   userid.jobname.facility.priority;userid.jobname.facility.priority
#
# When using syslogd rules with a series of conditions separated by
# semicolons, all of the individual conditions are evaluated
# left-to-right for each message. Each matching condition results in
# either a TRUE (meaning log the message) or a FALSE (meaning don't
# log the message). Conditions that don't match are ignored.
# The final result of evaluating each matching condition left-to-right
# is the result of the last matching condition. Rules that have no
# matching conditions for a message result in a FALSE.
# Matching exclude conditions (those with priority of none) result
# in a FALSE. As an example, consider the difference between the
# following two rules for a message with facility of daemon and
# a priority of emerg.
#
# daemon.none;*.emerg  /tmp/mylogfile
# *.emerg;daemon.none  /tmp/mylogfile
#
# The first rule, first condition, results in FALSE. The first rule,
# second condition, results in TRUE. Therefore, the message will be
# logged to the destination for this rule.
# The second rule, first condition, results in TRUE. The second rule,
# second condition, results in FALSE. The message will not be logged
# for this rule.
#
# The order of conditions within the filter is significant.
#
# Three criteria can be used to select messages received over the
# network for processing:
#
# 1) IP address or hostname of the sender. The IP address may be in
#  IPv4 or IPv6 format or may be a hostname that resolves to an
#  IPv4 or IPv6 address. If an IP address is used, an optional prefix
#  length may be specified with the /x notation.
#
# 2) facility of the message. See the description of facility above.
# 3) priority of the message. See the description of priority above.
#
# These criteria can be specified together as
#
#   (ip_address).facility.priority
#
#   or
#
#  (hostname).facility.priority
#
# If the the IP address or hostname is not to be considered in selecting the
# rule, then omit it and specify just facility.priority
#
# The following rule will match locally generated messages or
# messages received over the network from any source IP address
# that have the specified facility and priority (or higher).
#
# facility.priority
#
# The criteria for selecting messages for processing are combined
# with a destination, which tells syslogd what to do with selected
# messages.
#
#   criteria    destination
#
# The destination can be a file, one or more user IDs, SMF, syslogd
# at a remote host, or all logged-in users, or the operlog log stream.
#
# If the destination is a file, it may be optionally followed by two
# options, -F and -D. -F should be followed by an octal number that
# indicates the permissions value to be used if syslogd must create
# the file. -D should be followed by an octal number that indicates
# the permissions value to be used if syslogd must create the
# directory to contain the file. These options are only effective
# if syslogd is started with the -c start option. See the
# Communications Server IP Configuration Reference for details.
#
# The following example stores messages with facility daemon or
# local1 in the file /directory/logfile.
#
#   daemon.*;local1.*   /directory/logfile
#
# The directory structure used in this sample configuration is
# expected to be created automatically by syslogd, with a new
# directory of log files for each day.  This requires two types
# of configurations outside of the scope of this configuration
```

```
# file:
#
# 1) syslogd command-line option
#
#    The syslogd -c command-line option should be enabled, causing
#    syslogd to create log files and directories if they do not
#    already exist.
#
# 2) cron job
#
#    A cron job should be utilized to wake up syslogd at the
#    beginning of each day to switch to new log files in a new
#    directory.  Here is the cron job definition:
#
#      1 0 * * * kill -HUP `cat /etc/syslog.pid`
#
#    This job should be defined for a user ID with UID zero so that
#    it has permissions to send the signal to syslogd.
#
#    See UNIX System Services Planning and UNIX System Services
#    Command Reference for more information about cron.
#
# A sample shell script is provided for removing log files which are
# a specified number of days old.  It assumes the same directory
# structure which is used in this sample configuration.
#
# All example rules except for the last one are commented-out.  Some
# or all of the example rules will need to be changed for your
# environment.  Each example rule contains an explanation of changes
# which may be required.
#
##############################################################################
#
# Write all messages with priority crit or higher to the MVS operator
# console.  See the UNIX System Services Planning manual for more
# information about the /dev/console special file.
#
# *.crit             /dev/console
#
##############################################################################
#
# Write all messages with facility of daemon and a priority of error
# or higher to the operlog log stream. The operlog facility must be
# active in order to be able to log messages to the operlog log
# stream.
#
# daemon.err             /dev/operlog
##############################################################################
#
# Write all messages from syslogd itself to the file
# /var/log/YYYY/MM/DD/syslogd.log and to the system console.
#
# Notes:
#
# a) If syslogd is invoked as a started task or from a shell script
#    (e.g., /etc/rc) with job name SYSLOGD, the name of the
#    long-running syslogd job is SYSLOGD followed by a digit.
#
#    If syslogd runs with a different job name on your system, the
#    rule will have to be changed accordingly.
#
# b) During initialization, syslogd writes messages to
#    /dev/console.  These rules cover messages during steady-
#    state.
#
# *.SYSLOGD*.*.*      /var/log/%Y/%m/%d/syslogd
# *.SYSLOGD*.*.*      /dev/console
#
##############################################################################
#
# Write all messages from inetd to the log file inetd and to the
# console.
#
# Notes:
#
# a) If inetd is invoked as a started task or from a shell script
#    (e.g., /etc/rc) with job name INETD, the name of the
#    long-running inetd job is INETD followed by a digit.
#
#    If inetd runs with a different job name on your system, the rule
#    will have to be changed accordingly.
#
```

```
# *.INETD*.*.*       /var/log/%Y/%m/%d/inetd
# *.INETD*.*.*       /dev/console
#
#####################################################################
#
# Write all messages with priority err or higher from applications
# which specify facility "daemon" to the log file daemon.
# Because we chose to log messages from syslogd and inetd separately,
# we'll filter out those messages from this rule using special
# priority none.
#
# Notes:
#
# a) In this example, SYSLOGD followed by some other character is the
#    job name of syslogd.  If it is different on your system, change
#    the rule.
# b) In this example, INETD followed by some other character is the
#    job name of inetd.  If it is different on your system, change the
#    rule.
#
# daemon.err;*.SYSLOGD*.*.none;*.INETD*.*.none /var/log/%Y/%m/%d/daemon
#
#####################################################################
#
# Write all messages from applications which specify facility "auth"
# to the log file auth.
#
# auth.* /var/log/%Y/%m/%d/auth
#
#####################################################################
#
# Write all messages from applications which specify facility "mail"
# to the log file mail. Use file permissions of 640 octal if the file
# has to be created. Use permission of 770 octal if the directory has
# to be created. syslogd must be started with -c for these options
# to have any effect.
#
# mail.* /var/log/%Y/%m/%d/mail -F 640 -D 770
#
#####################################################################
#
# Write all messages with priority err and higher from otelnetd and
# other applications which specify facility "local1" to the log file
# local1.
#
# local1.err       /var/log/%Y/%m/%d/local1
#
#####################################################################
#
# Write all messages from otelnetd and other applications which
# specify facility "local1" when running as user SMITH to the log file
# local1.smith.  This could be useful if, for example, otelnetd traces
# need to be collected for a problem which user SMITH is experiencing
# and you do not wish to collect otelnetd traces from all user IDs.
#
# SmITh.*.local1.*  /var/log/%Y/%m/%d/local1.smith
#
#####################################################################
#
# Write all messages with priority err and higher to SMF.  These will
# be stored in SMF record type 109.  SMF must be active and
# configured to accept record type 109.  The user ID associated with
# syslogd must have read access to BPX.SMF.  See UNIX System Services
# Planning for more information about BPX.SMF.
#
# *.err        $SMF
#
#####################################################################
#
# Write all messages with priority crit and higher to the syslogd on
# host 192.168.1.9. The host may be specified by IPv4 address, by IPv6
# address, or by a name that resolves to an IPv4 or IPv6 address.
#
# *.crit      @192.168.1.9
#
#####################################################################
#
# Write all messages with priority crit and higher that arrive from
# host 192.168.0.6 to the operlog log stream.
#
# (192.168.0.6).*.crit      /dev/operlog
#
```

```
##################################################################################
#
# Write all messages with priority crit and higher that arrive from
# any host with IP address in the range 192.168.0.0 to 192.168.0.255
# to the operlog log stream.
#
# (192.168.0.6/24).*.crit      /dev/operlog
#
##################################################################################
#
# Write all messages with priority err and higher to log file errors.
#
# THIS EXAMPLE STATEMENT IS UNCOMMENTED.
#
*.err            /var/log/%Y/%m/%d/errors
#
```

Starting and stopping syslogd

You must start the syslog daemon (syslogd) from a user ID with superuser authority (UID 0). You can start syslogd from the z/OS UNIX System Services shell, from a started procedure using BPXBATCH, or from a started procedure that directly invokes syslogd. To ensure that syslogd services are available to applications during TCP/IP initialization, start syslogd from the z/OS UNIX System Services initialization shell script `/etc/rc`.

```
#
# Start the syslogd daemon
#    Export any syslogd environment variables before starting syslogd.
#
#    The ampersand (&) symbol is necessary to allow the z/OS Unix System
#    Services initialization to continue while syslogd starts.
#
#
#
export _BPX_JOBNAME='SYSLOGD1'
/usr/sbin/syslogd -f /etc/syslogd.conf &
```

If there is no TCP/IP transport active when syslogd starts or when TCP/IP is recycled, syslogd remains active until it can establish or reestablish communication with TCP/IP.

Rules:

- If you start syslogd from a shell script, start it as a background shell command by specifying an ampersand (&) as the last character on the command. If you do not specify the ampersand, control does not return to the shell until syslogd ends. When starting syslogd from the `/etc/rc` script it is especially important to code the ampersand; if you do not code the ampersand, z/OS UNIX System Services initialization cannot complete.
- If you start syslogd from a shell script and you call the shell script from a started procedure using BPXBATCH, you need to add a **sleep** shell command to your script after the start of syslogd. The **sleep** command prevents the shell script from ending before syslogd initialization completes.
- If you start syslogd from a shell script and you call the shell script from a started procedure using BPXBATCH, you must use a z/OS UNIX file for the STDOUT and STDERR DD statements; otherwise the started procedure cannot end.

Guidelines:

- When starting syslogd from a shell script, the resulting job name is the user ID under which the shell script is running. You can override the default job name by setting and exporting the _BPX_JOBNAME environment variable in your shell script.
- When starting syslogd from a started procedure, the resulting job name is the same as the name of the started procedure. You can override the default job name by setting the JOBNAME= parameter. For example:

  ```
  S SYSLOGD1,JOBNAME=SYSLOGD
  ```

- When starting syslogd from a shell script, export the environment variables before starting syslogd. The following example defines the syslogd configuration file:

```
#
# Shell script to start syslogd
#
export _BPX_JOBNAME='SYSLOGD1'
export SYSLOGD_CONFIG_FILE="//'HLQ.SYSLOGD.CONFIG(DEFAULT)'"
 /usr/sbin/syslogd &
```

- When starting syslogd directly from a started procedure, place the syslogd environment variables in a z/OS UNIX file or MVS data set. Use the following technique to pass the environment variables to syslogd.

```
// PARM='ENVAR("_CEE_ENVFILE=DD:STDENV")/'

//STDENV DD PATH='/etc/syslogd.env',PATHOPTS=(ORDONLY)

or

//STDENV DD DSN=HLQ.SYSLOGD.ENV(DEFAULT),DISP=SHR
```

  When you use an MVS data set for your syslogd environment variables, place the environment variables in a data set with the VB record format [RECFM(VB)] and a logical record length of 256 [LRECL(256)]. If you use any other record format for the data set, use the _CEE_ENVFILE_S environment variable in place of the _CEE_ENVFILE environment variable in your syslogd started procedure. When the _CEE_ENVFILE_S environment variable is used, the system removes trailing blank spaces from each *NAME=VALUE* line that is read. For additional information about the _CEE_ENVFILE_S environment variable, see z/OS XL C/C++ Programming Guide.

Syslogd can run in one of three modes:

- Normal

  In this mode, syslogd processes logging requests from the local system and applications using the syslog() function. Additionally, syslogd receives and logs messages sent over the IP network by remote systems running syslogd. These remote systems can be z/OS systems or non-z/OS systems. Only one instance of syslogd can be run on a z/OS system if syslogd is started in this mode.

- Local-only

  In this mode, syslogd processes logging requests from only the local system and applications. This instance of syslogd does not receive or process messages sent over the network from remote syslog daemons. Use the `-i` option to start syslogd in the local-only mode.

- Network-only

  In this mode, syslogd processes only messages sent over the network by remote systems running syslogd. This instance of syslogd does not process logging requests from the local system or applications. Use the `-n` option to start syslogd in the network-only mode.

In all three modes, syslogd can send messages to remote syslog daemons.

Requirement: You must install and activate the AF_UNIX domain prior to starting syslogd in normal or local-only mode.

As shown in Table 15 on page 220, how you choose to run syslogd depends on your logging requirements.

Table 15. Mode to use for different logging requirements

Logging requirement	**How to run syslogd**
You have a low volume of messages that you want to process from the network.	Run a single instance of syslogd in normal mode.
You do not want to process any log messages from remote syslog daemons over the network.	Run a single instance of syslogd in local-only mode.

Table 15. Mode to use for different logging requirements (continued)	
Logging requirement	**How to run syslogd**
You have a high volume of log messages from remote syslogd daemons that you want to log with syslogd.	Run two instances of syslogd. Start syslogd in local-only mode to process all local messages. Start another instance of syslogd in network-only mode to process the remote syslog messages. Two instances of syslogd provide better performance than running a single instance of syslogd in normal mode, especially in high-volume situations.

Restriction: A maximum of two instances of syslogd can be started. If you are going to start more than one instance of syslogd on the same z/OS image, one instance must be started in local-only mode and one instance must be started in network-only mode. Never run just one instance of syslogd in network-only mode. If an instance of syslogd is not processing local system and application messages, these messages are written to the MVS console and might result in message flooding on the MVS console.

To record timestamps more accurately, you need to set the TZ environment variable to local time. You can set the TZ environment variable in the following ways:

- When you start syslogd from the z/OS shell

  Export the TZ environment variable before you start syslogd. Export it in /etc/profile or in .profile in the HOME directory. For example, if you are in the Eastern time zone in the United States, issue the following command:

  ```
  export TZ=EST5EDT
  ```

- When you are starting syslogd as a started task, use either of the following methods:
  - Specify TZ using the ENVAR parameter on the PARM statement in the started procedure. For example:

    ```
    // PARM='ENVAR("TZ=EST5EDT")/'
    ```

  - Export the TZ environment variable in a file specified with the STDENV DD statement. For example:

    ```
    // PARM='ENVAR("_CEE_ENVFILE=DD:STDENV")/'
    //STDENV DD PATH='/etc/syslogd.env',PATHOPTS=(ORDONLY)
    ```

    Place the following statement in the /etc/syslogd.env file:

    ```
    TZ=EST5EDT
    ```

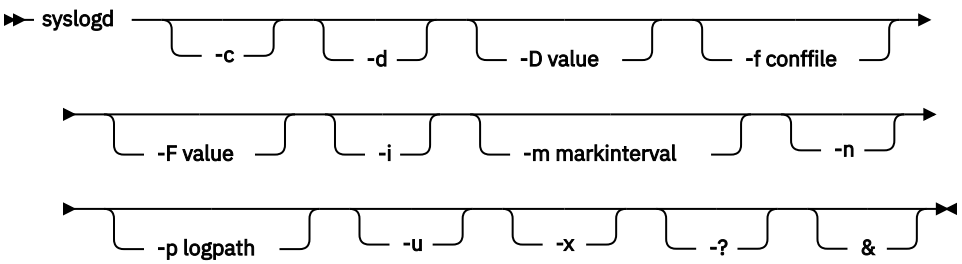
    Use the STDENV DD statement when you want to specify more than one environment variable; there is a JCL limit of 100 characters on the PARM parameter. Language Environment recommends a variable record format for the STDENV file.

You can also set the TZ environment variable for all applications in the CEEPRM*xx* PARMLIB member. You should define the TZ environment variable for all three LE option sets (CEEDOPT, CEECOPT, and CELQDOPT). For example:

```
CEECOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
CEEDOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
CELQDOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
```

For more information about specifying run-time options, see z/OS Language Environment Programming Guide. For details on setting the TZ environment variable, see z/OS UNIX System Services Command Reference.

The following syntax is the syntax for the syslogd command:



The **syslogd** command recognizes the following options:

-c
Create log files and directories automatically. The -c option is required to use the automatic archive function (see "Configuring syslogd for automatic archiving" on page 230).

-d
Run syslogd in debugging mode (see "Diagnosing syslogd configuration problems" on page 233 for more information).

-D
Default access permissions (modes) to be used by syslogd when creating directories. This parameter is valid only when specified in conjunction with the -c option. The parameter value is specified as an octal number 1- 4 characters in length. Leading zeros can be omitted. The following values can be ORed together to form the parameter value:

Value
Description

2000
Set GID

1000
Sticky bit (deletion restricted to owner or superuser)

0400
User read

0200
User write

0100
User list directory

0040
Group read

0020
Group write

0010
Group list directory

0004
Other read

0002
Other write

0001
Other list directory

If the -D option is not specified, the default value 0700 is used. Value 0000 and bits other than the ones shown are not valid. For example, you cannot set the set-UID bit for a directory. z/OS does not use the set-GID bit during directory creation regardless of whether the bit was set in the directory permissions.

-f

Configuration file name. You can also specify the configuration file using the SYSLOGD_CONFIG_FILE environment variable. The `-f` option overrides the environment variable.

-F

Default access permissions (modes) to be used by syslogd when creating log files. This parameter is valid only when specified in conjunction with the `-c` option. The parameter value is specified as an octal number 1 - 4 characters in length. Leading zeros can be omitted. The following values can be ORed together to form the parameter value:

Value
Description

0400
User read

0200
User write

0040
Group read

0020
Group write

0004
Other read

0002
Other write

If the -F option is not specified, the default value 600 is used. The actual permissions are modified by the syslogd process umask value at the time when the file is created. This parameter is used only when syslogd must create a log file dynamically; it has no effect on log files that already exist. Value 0000 and bits other than the ones shown are not valid. For example, you cannot set the execute bits, set-UID, set-GID, or the sticky bit for log files.

-i

Start in local-only mode, and do not receive messages from the IP network. This option is mutually exclusive with the `-n` option. If syslogd is started with the `-i` option, another instance of syslogd can be started with the `-n` option. This is the only supported way to run two instances of syslogd on the same z/OS image. Syslogd can still send messages to remote syslogd instances when it is running in local-only mode.

-m

Number of minutes between mark messages. The default value is 20 minutes. The following rule must be coded for each logfile that you want a mark record recorded in: mark.info.

-n

Start in network-only mode, and receive messages from only the IP network. This option is mutually exclusive with the `-i` option. If syslogd is started with the `-n` option, another instance of syslogd can be started with the `-i` option. This is the only supported way to run two instances of syslogd on the same z/OS image. Syslogd can still send messages to remote syslogd instances when it is running in network-only mode.

-p

Path name of the z/OS UNIX character device for the datagram socket. The default value is /dev/log. You can also specify the path name using the SYSLOGD_PATH_NAME environment variable. The `-p` option overrides the environment variable.

Note: This option is not used frequently. If you selected the `-p` option, syslogd will not function properly.

-u

For records received over the AF_UNIX socket (most messages generated on the local system), include the user ID and job name in the record. In this case, a forward slash, the user ID, and the job name will follow the local host name for messages received over the AF_UNIX socket. The forward

slash, which immediately follows the local host name, can be used to determine whether or not the user ID and job name is being recorded. If not recorded, a blank immediately follows the local host name. When user ID or job name is not available, N/A will be written in the corresponding field.

-x
Disable host name resolution for messages received from the IP network. This option is mutually exclusive with the `-i` option. Using this option can improve the performance of syslogd when processing messages received from the IP network. It has no effect for local messages. When you use this option, the IP address (instead of the host name) of the origin host is logged, along with the message text. If the host name can be determined from the rule without having to make a resolver call, the host name is used instead of the IP address. When the `-x` option is not used, syslogd always attempts to resolve the host name associated with a log message arriving from the IP network. If the host name cannot be determined, the IP address is logged as the message origin instead of the host name.

-?
Show syslogd command-line options.

To specify the job name and pass the appropriate environment variables to the syslogd process, start syslogd by using a shell script such as the following script:

```
#
# Start the syslog daemon
#

export _BPX_JOBNAME='syslogd'
/usr/sbin/syslogd -f /etc/syslog.conf &
```

You can execute this shell script directly from the /etc/rc file to start syslogd at z/OS UNIX initialization.

If an incorrect argument or number of arguments is entered, syslogd exits and the return code is 1. In all other situations in which syslogd exits, the return code is 0.

To terminate syslogd, you can issue the STOP command, issue the MODIFY command, or send a SIGTERM signal.

```
STOP jobname

MODIFY BPXOINIT,TERM=processID

kill -s TERM processID
```

To force syslogd to reread its configuration file and activate any modified parameters without stopping, issue the MODIFY *procname*,RESTART command or send a SIGHUP signal. This also causes any host names specified in rule selectors or destinations to be resolved again to IP addresses. After rereading the configuration file, syslogd continues to append log messages to the files that you specify in /etc/syslog.conf.

```
MODIFY procname,RESTART

kill -s HUP processID
```

The syslog daemon stores its process ID in the /etc/syslog.pid file or the /etc/syslog_net.pid file, so that it can be used to terminate or reconfigure the daemon. If syslogd is started in normal or local-only mode, the /etc/syslog.pid file is used to store the process ID. If syslogd is started in the network-only mode, the /etc/syslog_net.pid file is used to store the process ID.

Restrictions:

- If /etc/syslog.pid or /etc/syslog net.pid is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the syslog daemon.
- If /etc/syslog.pid or /etc/syslog_net.pid is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /etc/syslog.pid or /etc/syslog_net.pid is stored cannot have

write access to the directory. Additionally, write access to /etc/syslog.pid and /etc/syslog_net.pid must be limited to the owning UID or group, for example, --w--w----permissions.

Rule: If the BPX.SMF FACILITY class resource is defined and SMF records are to be written by syslogd, the user ID with which syslogd runs must also be permitted to SAF resource BPX.SMF. See SEZAINST(EZARACF) for more information.

Tips:

- Messages are read from the UNIX domain datagram socket (unless the `-n` option is specified), and from the IPv4 (AF_INET) or IPv6 (AF_INET6) Internet domain datagram socket (unless the `-i` option is specified).
- Messages written to a local instance of syslogd with the kern facility are converted to the user facility. If syslogd receives a log message over the network with the kern facility, the facility is not changed.

For more information about the facilities used by z/OS Communications Server functions, see Table 4 on page 30.

Configuring syslogd to receive remote messages

The ability to receive messages from remote syslogd instances can be very useful in providing a consolidated log of messages from multiple hosts into a consolidated z/OS message log. For example, in scenarios where you are running Linux hosts on System z processors, and these Linux hosts are performing processing in cooperation with or on behalf of z/OS systems, you might want to have certain important messages generated on the Linux hosts visible from a z/OS system. This capability would enable z/OS operators to be alerted of specific conditions on the Linux hosts that might require actions to be taken locally or on the Linux hosts.

If you decide that this remote logging capability is useful in your environment, there are several configuration considerations that should be examined prior to enabling this function. It is also important to note that the syslogd remote logging capability can work in both directions; the z/OS syslogd can forward some of its messages to another remote syslogd instance, or the z/OS syslogd instance can be the receiver of remote syslogd messages. The considerations described here focus primarily on the latter scenario, where the z/OS syslogd is the recipient of remote messages.

Improving the efficiency of syslogd remote logging functions

While the ability for the z/OS syslogd to receive remote messages can be very useful, it is important that appropriate planning take place to ensure that the additional remote message traffic that syslogd receives does not create a performance issue or impede the ability for the local z/OS syslogd to perform its processing. The following configuration options can help reduce such risks:

- Configure the remote syslogd instances to forward only messages that are important enough to consolidate on the z/OS system. For example, high-volume debug traces should be collected on the local host instead of having them forwarded to a z/OS system. Syslog daemon implementations typically provide the ability to define configuration statements that dictate which type of messages are forwarded to remote hosts. For example, the z/OS syslogd enables you to filter messages based on the facility and priority associated with a message. For more information about the specific configuration of your remote syslogd instances, consult the documentation available for the platform on which the remote syslogd instance is running.
- If the remote message traffic that the local z/OS syslogd instance is expected to receive is substantial, consider configuring a network-only instance of syslogd and a local-only instance of syslogd. This helps ensure that the logging of locally generated z/OS syslogd messages is not impacted by high-volume remote message traffic.
- When the local z/OS syslogd instance receives a message from a remote syslogd, it determines the disposition of the message based on its own configuration statements. These configuration statements enable you to determine which received messages are of interest and how these messages should be logged. The following guidelines describe how these statements can be configured for remote message receipt:

– The z/OS syslogd can log messages to various destinations, including z/OS File System files, the MVS console, the MVS operations log, SMF records, and so on. If the destination for the remote messages is the MVS console or MVS system log designated with the /dev/console destination, consider whether logging these messages to the MVS operations log (that is, the OPERLOG log stream) designated by the /dev/operlog destination is a suitable alternative. Logging messages to the OPERLOG log stream is more efficient and consumes less system resources than logging messages to the MVS console.

  The OPERLOG log stream must be configured and active before using the syslogd /dev/operlog destination. If the OPERLOG log stream is not active when syslogd attempts to log a message to the /dev/operlog destination, message FSUM1234 is logged with a facility.priority value of daemon.error. If the OPERLOG log stream later becomes active, message FSUM1235 is logged with a facility.priority value of daemon.info, and logging to OPERLOG automatically begins. For more information about using OPERLOG, see z/OS MVS Planning: Operations.

– You can identify which messages the z/OS syslogd should process by identifying the known remote syslogd instances from which you expect to receive messages. For example:

```
(host1.xyz.com).*.*        /dev/operlog
(host2.xyz.com).*.*        /dev/operlog
```

  In this example, syslogd resolves the specified host names to the IP addresses associated with those hosts during initialization by invoking the system resolver services. After initialization is complete and syslogd begins receiving remote messages, it first checks to determine whether the incoming messages match any of the specified configuration statements. In this example, syslogd processes only remote messages that have a source IP address associated with the hosts host1.xyz.com or host2.xyz.com, and logs these messages in the MVS operations log. Assuming that no other configuration statements are specified, any messages received from other hosts are discarded.

  Specifying configuration statements that identify each remote host using a host name also has the additional benefit of enabling these remote messages to include a host name identification when they are logged locally, without incurring the overhead of a host name resolver lookup for each incoming message. The per-message host name resolver lookup can be disabled using the `-x` syslogd option. If the ability to perform a resolver lookup for every message is necessary or useful in your environment, you should use the system resolver cache function (see “Customizing the resolver” on page 738 and “Resolver caching” on page 750). If host name resolution is performed, and the message was received over a link-local IP address, the resolved host name might include scope information that identifies the interface over which the message was received. For more details on support for scope information on a host name, see z/OS Communications Server: IPv6 Network and Application Design Guide.

  In addition to host names, remote syslogd instances can also be identified using IPv4 or IPv6 addresses or by specifying an IPv4 subnetwork or IPv6 network prefix. For example:

```
(10.42.105/24).*.*         /dev/operlog
(10.43.110.15).*.*         /tmp/otherlog
```

  In this example, any remote messages received from hosts using an IP address in subnetworks 10.42.105.0 to 10.42.105.255 are logged in the MVS operations log, while messages received from the host identified by 10.43.110.15 are logged in the /tmp/otherlog file.

– In addition to identifying the remote syslogd instances that can forward messages, you can also use additional filters to determine which remote messages you want to log locally. For example:

```
(host1.xyz.com).*.crit         /dev/operlog
(host2.xyz.com).*.crit         /dev/operlog
```

  With these configuration statements, syslogd logs only messages received from hosts host1 and host2 that have a critical priority to the MVS operations log. Assuming that no other configuration statements are specified, any other messages received from those hosts are discarded.

Security considerations

As with all network communication protocols, give careful consideration to the security requirements that might be appropriate to protect this remote message logging function in your environment. The following considerations are some to be examined:

- How do you ensure that the messages sent by a given host did in fact originate on that host, and that these messages have not been modified or replayed by a third party in the network?

  One mechanism that can be used to answer these questions is to deploy a Virtual Private Network (VPN) using IPSec across the remote syslogd instances and the local z/OS syslogd. IPSec configured with the Authentication Header (AH) protocol provides data integrity, data origin authentication, and an optional replay protection service. To deploy IPSec for syslogd traffic, configuration is required on both the local z/OS system and the remote hosts of the syslogd instances that are forwarding messages. On the z/OS side, this is accomplished by defining an IP security policy. In this IP security policy, you specify filter rules that indicate all the remote hosts that are allowed to send UDP traffic to the local syslogd on port 514, along with the IPSec actions that define the IPSec attributes for this traffic. The IPSec policy can be defined in such a way that if any UDP datagrams destined for local UDP port 514 are received that do not match the IP security policy, they are discarded by the local TCP/IP stack. This provides an additional level of protection against denial of service attacks for syslogd, as unauthorized messages are discarded without syslogd needing to process them.
- What if the forwarded messages contain sensitive data? How do you protect the confidentiality of the data as it traverses the network?

  You can use the IPSec Encapsulating Security Payload (ESP) protocol to provide data confidentiality (encryption). Note that the ESP protocol can also be used to provide some of the same functions that the AH protocol provides, such as data integrity, data origin authentication, and replay protection.

Obviously, the security considerations for deploying remote syslogd logging depend on your local environment and your local security policies. For more information about the z/OS configuration for IPSec, see “Overview of using IP security” on page 888. For more information about configuring IPSec on the remote hosts, consult the documentation for that platform.

Availability considerations

Syslogd availability is important to the logging of both local and remote messages. The following configuration considerations can help improve the availability of the z/OS syslogd remote logging services:

- Ensure that there is a process in place to restart a z/OS syslogd after a failure that results in the termination of syslogd. This can be accomplished by placing syslogd in the AUTOLOG list in the TCP/IP profile. This enables TCP/IP to initially start the syslogd instance as a started task, and also enables TCP/IP to monitor whether syslogd has a socket bound to the syslogd port used for remote message receipt (UDP port 514). This approach works well for single-stack INET configurations. If multiple TCP/IP stacks are deployed on a single system (that is, a CINET configuration), you should not use the AUTOLOG list. Alternatively, an installation can use any available automated operations software package to automate the start and restart of syslogd. For more information about the AUTOLOG statement, see z/OS Communications Server: IP Configuration Reference.
- On MVS systems that are not part of a sysplex and that have TCP/IP stacks with multiple network interfaces, using a static VIPA address can provide for better availability characteristics should a specific network interface or network experience an outage. This involves configuring a static VIPA in the TCP/IP profile or reusing an existing one. No special configuration of the local syslogd is needed. However, you should configure the remote syslogd instances to use the static VIPA address as the destination address when forwarding messages. This can typically be accomplished by either specifying the static VIPA as the destination address, or specifying a host name that maps to the static VIPA in the remote syslogd configuration.
- When the recipient syslogd instance is running in a sysplex environment, additional availability features are available that should be explored. For example, you can use a multiple application-instance dynamic VIPA (DVIPA) to represent the syslogd instance on a given system, and the remote syslogd

instances are configured to use that DVIPA address as the destination IP address for the messages they forward. In this configuration, if a failure to the MVS system or TCP/IP stack on which the syslogd instance is running occurs, the DVIPA is automatically moved to a predefined backup system that is currently active. This enables the syslogd instance on that backup system to begin processing these remote messages in a transparent manner. In this configuration, using the MVS operations log (OPERLOG) as the destination provides additional benefits, because the operations log is implemented as a coupling facility log stream that any system in the sysplex can access. For additional information related to static and dynamic VIPAs, see Chapter 7, "Virtual IP Addressing," on page 361.

Additional considerations

Syslogd on z/OS receives remote messages as UDP packets on port 514. Because UDP is a connectionless protocol that does not provide reliable communications, the potential exists that some UDP packets containing forwarded syslogd messages can be dropped, either in the network as a result of network congestion or on the z/OS system if the syslogd is overloaded and cannot process incoming messages fast enough. If this occurs, these messages are not seen by the z/OS syslogd instance, nor is the remote syslogd instance made aware of the dropped messages. As a result, make provisions to enable direct access of syslogd messages on the remote hosts. In addition, you should perform a detailed evaluation on the risks associated with the potential for message loss, prior to using any z/OS-based automated operations software to trigger actions based on the remote messages received by the z/OS syslogd.

Offloading log files

z/OS Communications Server includes a syslogd configuration file in /usr/lpp/tcpip/samples/syslog.conf, a REXX program for removing old log files in /usr/lpp/tcpip/samples/rmoldlogs, and a JCL procedure for starting syslogd in SEZAINST(syslogd). These are intended to be used together, though each may need to be customized for your installation.

Guideline: You should use this method of offloading or archiving files only if you do not use the automatic archive function of syslogd that is described in "Configuring syslogd for automatic archiving" on page 230. Because both of these methods rely on creating new log files, results are unpredictable if you try to use both methods together.

The sample syslogd configuration file is installed in /usr/lpp/tcpip/samples/syslog.conf. It can be copied to /etc/syslog.conf after customization. If it is copied somewhere else, the syslogd `-f` command-line option must be used to tell syslogd where to find the configuration file.

The sample REXX program for removing old log files is installed in /usr/lpp/tcpip/samples/rmoldlogs. It can be copied to an installation-defined directory after customization. The sample JCL procedure can be copied to an installation-defined library after customization.

The sample configuration uses date stamps in the names of directories of log files to organize log files by year (%Y), month (%m), and day (%d) as follows:

```
*.err      /var/log/%Y/%m/%d/errors
```

Log files for February 14, 2001, for example, would be stored in directory /var/log/2001/02/14. Variable substitution occurs using the Language Environment C function strftime(). Variables are case sensitive. For more information and a complete list of variables, see z/OS XL C/C++ Runtime Library Reference.

A cron job should be used to send the SIGHUP signal to syslogd every day at midnight so that it switches to a new set of files. The cron job should be created for a user ID with UID 0. The definition of the cron job is:

```
0 0 * * * kill -HUP `cat /etc/syslog.pid`
```

Tip: An accent mark (`) is used in this definition, not a single quotation mark.

The log file names vary based on the day, so sending SIGHUP to syslogd after the day changes causes syslogd to create new files.

Because some messages sent just after midnight may be logged by syslogd before it processes the SIGHUP signal, it is possible that a few messages sent after midnight will be stored in the log files for the previous day.

The sample REXX program can be run daily to remove all log files older than the number of days specified in the program. Comments in the REXX program describe how to configure the number of days. The definition of a cron job to run the REXX program every day at 1:00 A.M. is:

```
0 1 * * * localdir/rmoldlogs
```

localdir is the name of the installation-defined directory where the customized version of /usr/lpp/tcpip/samples/rmoldlogs was copied.

Setting permissions for log files and directories

When you specify the -c start option, syslogd creates log files and directories dynamically. By default, directories are created with the permissions value 0700, which means that only the owner can read, write, and list the contents of the directory. Similarly, if syslogd needs to create a file, the default permissions value is 600, which again means that only the owner can read and write to the file. Because a user ID with UID 0 must run syslogd, the owner is always a superuser. To change the default permissions used by syslogd, use either the -F or the -D start option to set the global default permissions for files and directories, respectively.

Tip: The -F and -D start options have no effect on files or directories that already exist.

You can also use the -F and -D configuration options to override global defaults for individual syslogd rules. Specify -F or -D (or both) with octal values following the file name. For example:

```
*.err      /var/log/%Y/%m/%d/errors -F 640 -D 644
```

The file permission bits, whether provided on the rule or as global defaults, are modified by the syslogd process file creation mask (umask), and then used to set the file permission bits of a file that is being created.

If you are considering allowing users other than a superuser to have access to log files, before changing the syslogd default permissions for files and directories, be sure to consider the following options:

- Before starting syslogd, create the log file (and containing directory if necessary) with permissions and ownership that allows the other users to have access. If a single user needs access, you can make the file user ID (UID) match that of the user ID that needs access. If multiple users need access, set a new or existing group ID (GID) as the file's GID, and set the permissions to allow members of the group to have read access, write access, or both. The file or directory UID and GID can be set with the **chown** command. Be sure to give the syslogd user ID write access to the log files. This technique is useful only if the files are not being created dynamically by syslogd.
- If you are not using file access control lists (ACLs), files and directories created by syslogd have the owner UID 0. By default, the owning GID is set to that of the parent directory. However, if the FILE.GROUPOWNER.SETGID profile exists in the UNIXPRIV class, the owning GID is determined by the set-GID bit of the parent directory, as follows:
  - If the set-GID bit of the parent directory is on, the owning GID is set to that of the parent directory.
  - If the set-GID bit of the parent directory is off, the owning GID is set to the effective GID of the process.

  When there are no file access control lists, the only way to manage log files with different access requirements that must be accessed by different groups of users is to create the containing directories with the appropriate GIDs before starting syslogd, and let syslogd dynamically create the log files in the appropriate directories. The log files then inherit the GID of the directory, if the directory has the set-GID bit on.
- A third way to provide access to log files for different users or groups of users is to use file access control lists. For information about setting file access control lists, see the **setfacl** command in z/OS UNIX System Services Command Reference. The ACLs for dynamically created directories and files can

be inherited from defaults set on the parent directory. When using this method, be sure that the syslogd user ID continues to have write access to the log files.

Configuring syslogd for automatic archiving

You can set up syslogd to perform automatic archival of eligible z/OS UNIX files that are configured as destinations on syslogd rules. Eligible files are those that you tag using the -N or -X parameter on the rule definitions. You can choose to archive at a specific local time of day, or when the file systems that contain the eligible files become too full, or both. You can also perform an on-demand archive using an operator command.

Guideline: You should use this method of archiving files only if you do not offload files using the provided sample configuration and procedure that are described in "Offloading log files" on page 228. Because both of these methods rely on creating new log files, results are unpredictable if you try to use both methods together.

To configure syslogd for automatic archiving, take the following steps:

- Configure the events that trigger automatic archival
- Configure the archive details for each z/OS UNIX file

Steps for configuring the events that trigger automatic archival

You can set up syslogd to perform automatic archival of eligible z/OS UNIX files that are configured as destinations on syslogd rules. Use this method of archiving files only if you do not offload log files.

Before you begin

Automatic archiving can be performed at a specific local time every day, or when one or more of the z/OS UNIX file systems that contain the eligible files become too full. Determine whether you want to configure one or both of these triggers.

Requirement: You must specify the `-c` start option when you are using the automatic archive function.

Procedure

Perform the following steps to configure syslogd archive triggers:

1. Specify the ArchiveTimeOfDay statement in the syslogd configuration file to specify the local time of day using hours and minutes in a 24 hour clock format.
2. Specify the ArchiveThreshold statement in the syslogd configuration file to specify the percentage of the file system that must be full to trigger an automatic archive.

   Guideline: You should set up the syslogd file destinations as one or more separate z/OS UNIX file systems. Threshold-based archiving works best when the file systems contain only data written by syslogd.
3. Specify the ArchiveCheckInterval statement in the syslogd configuration file to specify the interval in minutes at which syslogd should check to determine how full the file systems are.

Results

The configuration statements that define archive triggers are global statements. If you configure these statements multiple times, the last instance is used.

Steps for configuring the archive details for each z/OS UNIX file

Automatic archiving operates on eligible z/OS UNIX files. You can configure which files are eligible by specifying parameters on individual rules. You can also configure global parameters that affect the archive process and determine the destination data sets for each file.

Before you begin

You need to protect the syslogd archive data sets using available data set access controls to ensure that only authorized personnel are allowed to read these data sets. Syslogd archive data sets might contain sensitive information.

Procedure

Perform the following steps to configure the archive details for each z/OS UNIX file:

1. Specify the BeginArchiveParms statement in the syslogd configuration file.

   The BeginArchiveParms statement specifies a data set name prefix that is extended with a unique qualifier for each file to be archived, for the set of rules that follows this statement.
2. Specify one of the following parameters on each rule that uses a z/OS UNIX file to indicate that the archive function must process the z/OS UNIX file:

   - The -N parameter specifies that the file must be automatically archived and then re-initialized when an archive event occurs.
   - The -X parameter specifies that the file must be only re-initialized but not archived when an archive event occurs.

   If you do not specify the -N or -X parameters on a rule, then nothing happens to the destination file when an archive event occurs. See Supported destinations for syslogd in z/OS Communications Server: IP Configuration Reference for more information about specifying the -N or -X parameters.

   If you use generation data group (GDG) data sets as an archive destination, the GDG base must already be created. The following sample JCL creates a GDG base called USER1.SYSARCH.

```
//USER1X   JOB MSGLEVEL=(1,1),MSGCLASS=D,NOTIFY=USER1
//GDGA EXEC PGM=IDCAMS
//*
//GDGMOD   DD DSN=USER1.SYSARCH,
//      VOL=SER=CPDLB1,
//          UNIT=SYSALLDA,
//          SPACE=(TRK,(0)),
//          DCB=(LRECL=80,RECFM=FB,BLKSIZE=6800,DSORG=PS),
//          DISP=(,KEEP)
//SYSPRINT DD SYSOUT=*
//SYSIN    DD *
  DEFINE GENERATIONDATAGROUP -
         (NAME(USER77.MYGDG) -
         EMPTY -
         NOSCRATCH -
         LIMIT(255) )
```

   For more information about GDG data sets, see z/OS DFSMS Using Data Sets.

Using syslogd for z/OS UNIX application programs

You can use the logging facilities of the syslogd server with your z/OS UNIX application programs. Include the syslog.h header file with C programs so that they can open a log facility, send log messages to syslogd, and close the facility:

```
#include <syslog.h>

1 openlog("oec", LOG_PID, LOG_LOCAL0);
2 syslog(LOG_INFO, "Hello from oec");
3 closelog();
```

1 Open a log facility with the name of local0. Prefix each line in the log file with the program name (oec) and the process ID.

2 Log an info priority message with the specified content.

3 Close the log facility name.

The preceding statements created the following line in the log file:

```
May 26 11:27:51 mvs18oe oec[3014660]: Hello from oec
```

Restriction: The message length is limited to 4092 characters when written to a z/OS UNIX file. The message length includes any content prepended by syslogd. Any characters beyond the 4092 limit are discarded from the message.

For more information about the syslog function, see *Advanced Programming in the UNIX Environment*, published by Addison-Wesley or z/OS XL C/C++ Runtime Library Reference.

Usage notes

- It is possible to run two instances of syslogd. One instance must be started so that it processes messages from only the local host (`-i` option); the other instance must be started so that it processes messages from only the network (`-n` option).
- If you run two instances of syslogd, one for local messages and another for network messages, and you also configure the automatic archival function, do not configure the same UNIX file destinations in the two configuration files. The archival function renames, closes, and reopens the UNIX files. If two instances of syslogd are performing the archival function on the same set of files, results of the archival function are unpredictable. The same is true for the configured archive destination data sets. Be sure to configure unique UNIX file destinations and archive data set names for the two syslogd instances.
- syslogd can run swappable or nonswappable. When an application makes an address space nonswappable, it might convert additional real storage in the system to preferred storage. Because preferred storage cannot be configured offline, allowing syslogd to run in a nonswappable state can reduce the installation's ability to reconfigure storage in the future. Use the following guidelines to set the wanted state:
  - If the FACILITY class resource BPX.STOR.SWAP is not defined to the system:
    - syslogd will run nonswappable.
    - syslogd cannot be prevented from running nonswappable.
  - If the FACILITY class resource BPX.STOR.SWAP is defined to the system with UACC(NONE):
    - syslogd will run swappable by default (no access to BPX.STOR.SWAP).
    - syslogd can run nonswappable (given at least READ access to BPX.STOR.SWAP).
  - To define the FACILITY class resource BPX.STOR.SWAP issue the following commands:

    ```
    RDEFINE FACILITY BPX.STOR.SWAP UACC(NONE)
    SETROPTS RACLIST(FACILITY)REFRESH
    ```
- If you want syslogd to receive log data from remote syslogd servers, ensure that syslogd can bind to UDP port 514 by reserving that port for the syslogd job in your PROFILE.TCPIP data set. Ensure that the syslog service is defined in your services file or data set (for example, /etc/services). The following example port reservation in PROFILE.TCPIP assumes that syslogd runs as job syslogd1:

  ```
  PORT
  ...
  514 UDP syslogd1     ;syslogd daemon
  ...
  ```

  The following example shows the services file or data set file entry:

```
syslog     514/udp
```

- Configuration file errors are written to the operator console because initialization is not complete until the entire configuration file has been read.
- Facility mark is not affected by the *.priority usage. Mark messages are written only to the destinations of rules that specify mark.info.
- If a mark interval of zero minutes is specified, mark messages will be written every thirty seconds.

Diagnosing syslogd configuration problems

You must install and activate the AF_UNIX domain before starting syslogd in normal or local-only mode. To determine whether AF_UNIX was successfully started, check for a message in the console log similar to the following message:

```
BPXF203I DOMAIN AF_UNIX WAS SUCCESSFULLY ACTIVATED.
```

syslogd supports a debug mode, which is selected using the `-d` command-line option. If you run syslogd from the UNIX shell using this debug mode, syslogd writes debug messages to STDOUT. These messages can be used to diagnose problems in the syslogd configuration or to collect documentation when reporting a syslogd problem to IBM support.

You can use the SYSLOGD_DEBUG_LEVEL environment variable to limit the amount of debug messages. You can specify the following debug levels:

Value
Level

1
Base debugging information.

2
Configuration file processing.

4
Message handling information for messages being logged by syslogd.

8
Automatic archive processing.

16
Operator command processing.

32
Thread-specific processing.

64
Mutex lock processing. Locks that are specific to threads are logged only if the debug level includes 32.

You can add these values together in any combination to select the type of debug messages to be written. For example, SYSLOGD_DEBUG_LEVEL=91 includes all debugging information except for message handling and thread-specific processing (including locks). The default debug level is 127, which includes all debug information.

Rule: Do not use the `-d` option for normal operations.

If you are running syslogd in batch with `-d`, debug output is written to SYSPRINT, SYSTERM, or SYSERR, whichever is found first. The sample syslogd procedure SEZAINST(syslogd) defines SYSPRINT so that debug messages are stored in the job output. The following example shows how to change the SYSPRINT DD statement to write debug output to a file.

```
SYSPRINT DD PATH='/var/syslog/syserr',PATHOPTS=(OWRONLY,OCREAT)
```

Use caution using -d when syslogd is started from /etc/rc. If -d is used in this way, the shell background character (&) must be used to run syslogd in the background. Otherwise, /etc/rc does not end and UNIX System Services initialization does not complete.

If you are using the automatic archive function of syslogd, it is possible that an archive of a particular file fails. A failure can occur for many reasons, including dynamic allocation errors for the target data set, or read/write errors for the UNIX file or target data set. If a failure occurs, syslogd attempts to archive the file again when the next archive trigger event occurs, but it is possible that all archive attempts continue to fail. If all attempts fail, the original UNIX file to be archived still exists in the UNIX file system, renamed with a suffix of the form D*yymmdd*.T*hhmmss*, and the target data set might also exist and contain partial data. You should examine the error messages in the syslogd destination output file for the daemon facility to determine which files have failed, and manually recover or delete such files and the associated partial archive data sets.

Syslog daemon name/token pair and ECSA storage mapping

When the syslog daemon (syslogd) is started, a system-level name/token pair is created. The token name is EZBSYSLOGD, and the token value is the address of the ECSA storage that contains the syslogd configuration file location and other related information. For more information about the syslog daemon name/token pair and its ECSA storage mapping, see z/OS Communications Server: IP Programmer's Guide and Reference.

Configuring TCPIP.DATA

The TCPIP.DATA configuration data set is the anchor configuration data set for the TCP/IP stack and all TCP/IP servers and clients running in z/OS. With a z/OS TCP/IP stack, you can define the TCPIP.DATA parameters in a z/OS UNIX file or in an MVS data set. The TCPIP.DATA configuration data set is read during initialization of all TCP/IP server and client functions. All functions must access this data set in order to find basic configuration information, such as the name of the TCP/IP address space, the TCP/IP host name, and the data set prefix to use when searching for other configuration data sets. The TCPIP.DATA file contains the following major groups of configuration parameters:

- Application operating characteristics
- Resolver operating characteristics
- Socket library diagnostic data statements
- Resolver diagnostic data statements

Use of TCPIP.DATA and /etc/resolv.conf

The TCPIP.DATA data set is also known as one of the resolver configuration data sets. In fact, this name is now more commonly used to refer to this important file in the UNIX System Services environment because the socket library contains a component called the resolver. In a UNIX system, you use the /etc/resolv.conf file for the same purpose as you use TCPIP.DATA in your MVS system.

TCPIP.DATA specifies the name of the TCP/IP address space. Because the data set search order can vary, your installation will determine which data set you can use. See Chapter 2, “IP configuration overview,” on page 11 for search order, data set, and file retrieval information.

If you use TCPIP.DATA, it can be shared between multiple systems with a system name. But, if TCPIP.DATA is allocated by SYSTCPD DD and an application forks, any allocations from the parent of SYSTCPD are lost to the child process.

In z/OS UNIX System Services, each application that uses the XL C/C++ API can have its own environment variable, RESOLVER_CONFIG='*xxx*'. There are no concerns for forked child processes; however, this means that you cannot share the data set or file among multiple systems.

Creating TCPIP.DATA

Create a TCPIP.DATA file by copying the sample provided in SEZAINST(TCPDATA) and modifying it to suit your local conditions.

Allocate this data set with either sequential (PS) or partitioned (PO) organization, a fixed (F) or fixed block format (FB), a logical record length (LRECL) between 80 and 256, and any valid block size for a fixed block. This file can also be the file /etc/resolv.conf, or a z/OS UNIX file that is pointed to by either the environment variable RESOLVER_CONFIG or the SYSTCPD DD in a JCL procedure. If you have a z/OS UNIX file, the maximum line length can be 256. The environment variable RESOLVER_CONFIG can also point to an MVS data set or PDS.

You can use any name for the TCPIP.DATA data set if you access it using the //SYSTCPD DD statement, or use ENVAR to set RESOLVER_CONFIG, in the JCL for all the servers, logon procedures, and batch jobs that execute TCP/IP functions. If you are not using the //SYSTCPD DD statement, the environment variable, or /etc/resolv.conf, then the data set name must conform to the conventions described in "Configuration files for the TCP/IP stack" on page 23. Another alternative is to use the well-known data set name SYS1.TCPPARMS(TCPDATA). The HOMETEST command, with Trace Resolver active, can be used to verify the data set name that the system finds for TCPIP.DATA. However, because HOMETEST is an MVS sockets application, it does not use RESOLVER_CONFIG or /etc/resolv.conf in its search order. For this reason, it is recommended that /etc/resolv.conf and TCPIP.DATA contain exactly the same information or consider using the resolver GLOBALTCPIPDATA setup statement.

Rule: Because TCPIP.DATA statements might need to be read and used multiple times by the resolver, the FREE=CLOSE JCL parameter should not be used when allocating SYSTCPD. To allow TCPIP.DATA statements to be changed while still allocated for long running programs, consider using a member of an MVS partitioned data set instead of an MVS sequential data set. For these long running applications, the resolver MODIFY REFRESH command should then be used to indicate that TCPIP.DATA statements have been changed.

TCPIP.DATA statements

Each configuration statement can be preceded by an optional *system_name*. This permits configuration information for multiple systems to be specified in a single *hlq*.TCPIP.DATA data set. The *system_name* is matched against the name of the system on which you are running. The name of the system is specified as part of the VMCF subsystem initialization, as the *nodename* value on the following parameter and statement:

- P= start parameter of the EZAZSSI started procedure
- VMCF statement of the IEFSSN*xx* member of parmlib

For more information about starting VMCF, see "Step 3: Configure VMCF and TNF" on page 117.

The statements are processed in the order they appear in the data set. The following rules apply to this processing:

- If the *system_name* does not match the name of the system, the configuration statement is ignored.
- If *system_name* is blank, the configuration statement is in effect on every system.
- If the *system_name* matches the host's name, the configuration statement that follows it is in effect.
- The last statement that matches is effective.

For example, if you have the following three TCPIPJOBNAME statements, MVS6 would look for a TCP/IP cataloged procedure named TCPBTA2, MVSA would look for TCPV3, and all other systems would look for TCPMCWN.

```
      TCPIPJOBNAME TCPMCWN
MVS6: TCPIPJOBNAME TCPBTA2
MVSA: TCPIPJOBNAME TCPV3
```

But if you reversed the order, all systems would try to find the procedure named TCPMCWN.

```
MVS6: TCPIPJOBNAME TCPBTA2
MVSA: TCPIPJOBNAME TCPV3
      TCPIPJOBNAME TCPMCWN
```

Take special care with those TCPIP.DATA statements that can be specified multiple times, such as the SEARCH, SORTLIST, NSINTERADDR/NAMESERVER, OPTIONS, and LOADDBCSTABLES statements. Because these statements are cumulative, if you specify a *system_name* value, it might need to be used on all instances of the same statement. Consider the following examples.

Example 1:

```
                NSINTERADDR 5.5.5.5
                NSINTERADDR 1.1.1.1
           TN03: NSINTERADDR 1.1.1.1
           TN03: NSINTERADDR 2.2.2.2
           TN04: NSINTERADDR 3.3.3.3
           TN04: NSINTERADDR 4.4.4.4
```

After these TCPIP.DATA statements are processed, the following ordered list of DNS addresses is available:

```
On TN03:
     5.5.5.5
     1.1.1.1
     1.1.1.1
     2.2.2.2
On TN04:
     5.5.5.5
     1.1.1.1
     3.3.3.3
     4.4.4.4
All others:
     5.5.5.5
     1.1.1.1
```

Example 2:

```
           TN03: NSINTERADDR 1.1.1.1
           TN03: NSINTERADDR 2.2.2.2
           TN04: NSINTERADDR 3.3.3.3
           TN04: NSINTERADDR 4.4.4.4
                 NSINTERADDR 5.5.5.5
                 NSINTERADDR 1.1.1.1
```

After these TCPIP.DATA statements are processed, the following ordered list of DNS addresses is available:

```
On TN03:
     1.1.1.1
     2.2.2.2
     5.5.5.5
     1.1.1.1
On TN04:
     3.3.3.3
     4.4.4.4
     5.5.5.5
     1.1.1.1
All others:
     5.5.5.5
     1.1.1.1
```

A sample TCPIP.DATA data set (TCPDATA) can be found in SEZAINST. For detailed information on each of the statements, see z/OS Communications Server: IP Configuration Reference.

Using MVS system symbols in TCPIP.DATA

For ease of management when configuring a complex environment, use MVS system symbols such as &SYSCLONE, &SYSNAME, and &SYSPLEX. The resolver translates these symbols as it reads TCPIP.DATA, reducing the number of TCPIP.DATA files that must be maintained in a multisystem environment. For

detailed information about symbols and how to define them, see z/OS MVS Initialization and Tuning Reference.

Configuring PROFILE.TCPIP

During TCP/IP address space initialization, a configuration profile data set (PROFILE.TCPIP) is read that contains system operation and configuration parameters. A sample data set, SEZAINST(SAMPPROF), can be copied and modified for use as your default configuration profile.

If you are not familiar with the search order for this data set, see "PROFILE.TCPIP search order" on page 23 for information about understanding data set search orders. See z/OS Communications Server: IP Configuration Reference for the complete statement syntax and descriptions of the configuration statements.

For ease of management when configuring a complex environment, you can use one of the following PROFILE.TCPIP data set features:

- Group related statements into separate files and use the INCLUDE statement in PROFILE.TCPIP to include them in your configuration.
- Use MVS system symbols (such as &SYSCLONE, &SYSNAME, and &SYSPLEX). Because TCP/IP translates these symbols as it reads this file, this feature reduces the number of PROFILE.TCPIP data sets that must be maintained in a multi-TCP/IP environment.

Note: For detailed information about symbols and how to define them, see z/OS MVS Initialization and Tuning Reference.

The PROFILE data set contains the following major groups of configuration parameters:

- TCP/IP operating characteristics
- TCP/IP physical characteristics
- TCP/IP reserved port number definitions (application configuration)
- TCP/IP routing definitions
- TCP/IP diagnostic data statements

This information explains the first three areas of configuration. For routing configuration information, see Chapter 6, "Routing," on page 279. For information about configuring diagnostic statements, see z/OS Communications Server: IP Diagnosis Guide.

Changing configuration information

If you want to change the TCP/IP configuration without stopping and starting the TCP/IP address space, you can dynamically change many of the TCP/IP configuration options established by the PROFILE.TCPIP data set. To make dynamic changes, put the changed configuration statements in a separate data set and process it with the VARY TCPIP,,SYNTAXCHECK and VARY TCPIP,,OBEYFILE commands.

- VARY TCPIP,,SYNTAXCHECK

  The VARY TCPIP,,SYNTAXCHECK command checks the syntax of statements in the data set but does not apply any changes to the active configuration.
- VARY TCPIP,,OBEYFILE

  The VARY TCPIP,,OBEYFILE command checks the syntax of statements in the data set and also applies changes to the active configuration.

  Statements in a VARY TCPIP,,OBEYFILE data set that are coded with syntax errors might not be applied to the active configuration as expected; activating a profile data set that contains syntax errors can lead to unscheduled outages.

Guideline: Use the VARY TCPIP,,SYNTAXCHECK command to check for syntax errors in a profile data set before activating it as the initial profile or activating it with the VARY TCPIP,,OBEYFILE command.

Result: Changes that are applied to the active configuration with the VARY TCPIP,,OBEYFILE command remain in effect only until you stop the stack. You can retain these configuration changes by integrating them into the initial profile, or by using the VARY TCPIP,,OBEYFILE command each time you start TCP/IP.

For more information about VARY TCPIP commands, see z/OS Communications Server: IP System Administrator's Commands. For information about dynamically changing settings for a particular configuration statement, see the *Modifying* information under each configuration statement in z/OS Communications Server: IP Configuration Reference.

Guideline: If you attempt to edit the PROFILE.TCPIP data set while TCP/IP is active, and PROFILE.TCPIP is defined in the TCPIP PROC as a sequential data set (for example, //PROFILE DD DISP=SHR,DSNAME=TCPIP.PROFILE.TCPIP), the `Data set in use` message might be displayed. To avoid this situation, specify FREE=CLOSE, as shown in the following example:

```
//PROFILE  DD DISP=SHR,DSNAME=TCPIP.PROFILE.TCPIP,FREE=CLOSE
```

Specifying FREE=CLOSE enables you to edit the profile while TCP/IP is active. Typically, when TCP/IP starts, it keeps the profile allocated and does not release the allocation until the end of the step (in this case, the end of the job). When you specify FREE=CLOSE, the release occurs after the data set is read, and MVS releases the enqueue on the profile so that you can edit it.

If the profile is a member of a partitioned data set (PDS), such as SYS1.TCPPARMS(PROFILE), FREE=CLOSE is not needed.

Setting up TCP/IP operating characteristics in PROFILE.TCPIP

Figure 44 on page 240 shows a portion of the sample configuration file for the TCP/IP address space, PROFILE.TCPIP. This sample can be copied from SEZAINST(SAMPPROF). Figure 44 on page 240 includes the portion of the sample that shows how to set up TCP/IP operating characteristics. Descriptions for the statements follow Figure 44 on page 240.

```
; ====================================================================
; ====================================================================
; General TCP/IP address space configuration
; ====================================================================
;
; ARPAGE: Specifies the number of minutes between creation or
;   revalidation of an LCS ARP table entry and the deletion of the
;   entry.
;
ARPAGE 20
;
; --------------------------------------------------------------------
;
; GLOBALCONFIG: Provides settings for the entire TCP/IP stack
;
; Example GLOBALCONFIG to offload TCP segmentation to OSA-Express
; features
;
; GLOBALCONFIG SEGMENTATIONOFFLOAD
;
; Example GLOBALCONFIG to exploit HiperSockets multiple write
; support
;
; GLOBALCONFIG IQDMULTIWRITE
;
; Example GLOBALCONFIG to displace TCP/IP CPU cycles onto a zIIP
; for certain workloads
;
; GLOBALCONFIG ZIIP IPSECURITY IQDIOMULTIWRITE
;
; Example GLOBALCONFIG to assign OSA-Express QDIO write priority
; values to packets associated with WorkLoad Manager service classes,
; and to forwarded packets
;
; GLOBALCONFIG WLMPRIORITYQ
;              IOPRI1 0
;              IOPRI2 1
;              IOPRI3 2 3
;              IOPRI4 4 5 6 FWD
;
; Example GLOBALCONFIG to enable SMC-R and SMC-D processing
```

```
;
; GLOBALCONFIG SMCD FIXEDMEMORY 1000
;                             SMCR PFID 203 PFID 205 FIXEDMEMORY 1000
;
; Example GLOBALCONFIG to enable zERT processing
;
; GLOBALCONFIG ZERT AGGREGATION INTVAL 2 SYNCVAL 00:00
; ------------------------------------------------------------------
;
; IPCONFIG: Provides settings for the IPv4 IP layer of TCP/IP.
;
; Example IPCONFIG for single stack/single system:
;
IPCONFIG DATAGRAMFWD SYSPLEXROUTING
;
; Example IPCONFIG for automatic activation of inter-stack dynamic XCF
;   and Same Host (IUTSAMEH) interfaces
;
; IPCONFIG DYNAMICXCF 201.1.10.10 255.255.255.0 2
;
; Example IPCONFIG for IPSECURITY support:
;
; IPCONFIG IPSECURITY
;
; Example IPCONFIG to provide accelerated forwarding at the DLC layer
;   for OSA-Express QDIO and HiperSockets packets
;
; IPCONFIG QDIOACCELERATOR
;
; ------------------------------------------------------------------
;
; IPCONFIG6: Provides settings for the IPv6 IP layer of TCP/IP.
;
; Example IPCONFIG6 to enable IPv6 packet forwarding and the use of
;   virtual IP addresses as source addresses in outbound datagrams:
;
; IPCONFIG6 DATAGRAMFWD SOURCEVIPA
;
; Example IPCONFIG6 for automatic activation of inter-stack dynamic XCF
;   and Same Host (IUTSAMEH) interfaces
;
; IPCONFIG6 DYNAMICXCF 2001::151:0000
;
; ------------------------------------------------------------------
;
; SOMAXCONN: Specifies maximum length for the connection request queue
;   created by the socket call listen().
;
SOMAXCONN 10
;
; ------------------------------------------------------------------
;
; TCPCONFIG: Provides settings for the TCP layer of TCP/IP.
;            RESTRICTLOWPORTS limits access to ports below 1024
;            to authorized applications.  Applications can be
;            authorized to low ports in three ways:
;             - via PORT or PORTRANGE with the appropriate jobname
;                or wildcard jobname
;             - APF authorized
;             - superuser
;
TCPCONFIG TCPSENDBFRSIZE 32K TCPRCVBUFRSIZE 32K SENDGARBAGE FALSE
         RESTRICTLOWPORTS
;
; Example TCPCONFIG to change the KEEPALIVE interval for applications
; that enable the SO_KEEPALIVE socket option but do not override
; the interval using the TCP_KEEPALIVE socket option.
;
; TCPCONFIG INTERVAL 30
;
; Example TCPCONFIG for AT-TLS support:
;
; TCPCONFIG TTLS
;
; ------------------------------------------------------------------
;
; UDPCONFIG: Provides settings for the UDP layer of TCP/IP
;            RESTRICTLOWPORTS limits access to ports below 1024
;            to authorized applications.  Applications can be
;            authorized to low ports in three ways:
;             - via PORT or PORTRANGE with the appropriate jobname
;                or wildcard jobname
```

```
;             - APF authorized
;             - superuser
;
UDPCONFIG RESTRICTLOWPORTS
;
; ------------------------------------------------------------------------
;
; SRCIP: Provides the following functionality:
;   - Provides for the substitution of a source IP address on a
;     jobname-specific or destination-specific basis, for applications
;     which specify either the IPv4 INADDR_ANY address, or the IPv6
;     unspecified address (in6addr_any) for the source IP address.
;     This may be done when an application issues an explicit bind()
;     call with either of these addresses, or when it bypasses issuing
;     an explicit bind() call and issues a connect().
;   - Provides the ability to designate if default source address
;     selection should prefer a public or a temporary IPv6 address
;     for the specified jobs.
;
;
; Example SRCIP to substitute a source IP address
;
;SRCIP
; JOBNAME      USER15            9.43.242.5
; JOBNAME      USER*             9.43.242.4
; JOBNAME      USER15            2001::092B:F203
; JOBNAME      JOB*              ETHER1
; DESTINATION  9.67.114.02       9.43.240.7
; DESTINATION  2003::090C:F246   INTF1
; JOBNAME      *                 9.43.242.3
; JOBNAME      *                 9.43.242.3
; JOBNAME      PAYROLL*          9.42.242.5              BOTH
; JOBNAME      SERVER1           9.42.242.4              SERVER
; JOBNAME      CLIENT*           2001:0DB8::9:43:242:6   CLIENT
;ENDSRCIP
;
; Example SRCIP to cause default source address selection to prefer
; public or temporary IPv6 addresses
;
;SRCIP
; JOBNAME      IPV6PUB            PUBLICADDRS
; JOBNAME      IPV6TEMP           TEMPADDRS
;ENDSRCIP
;
; ------------------------------------------------------------------------
;
; DEFADDRTABLE: Can be used to configure the policy table for IPv6
; default address selection.
;
;DEFADDRTABLE
; Prefix          Precedence Label
; ::1/128           50         0
; ::/0              40         1
; 2002::/16         30         2
; ::/96             20         3
; ::ffff:0.0.0.0/96 10         4
;ENDDEFADDRTABLE
```

Figure 44. Example of TCP/IP operating characteristics in PROFILE.TCPIP

The following information describes the statements that are shown in Figure 44 on page 240. For more information about any of these statements, see z/OS Communications Server: IP Configuration Reference. For information specific to IPv6 support, see z/OS Communications Server: IPv6 Network and Application Design Guide.

ARPAGE
Use ARPAGE to set the number of minutes between a revalidation and deletion of ARP table entries for LCS devices. If you want to describe this value in seconds versus minutes, use the IPCONFIG ARPTO statement.

GLOBALCONFIG
Use GLOBALCONFIG to print several counters in text format. These counters include number of TCP retransmissions and total number of TCP segments sent from the TCP/IP system. Most installations use the SMF facility of MVS to collect these counters in a more standard way.

Use GLOBALCONFIG to enable use of Shared Memory Communications over RMDA (SMC-R) and Shared Memory Communication - Direct Memory Access (SMC-D). For more details about SMC-R and SMC-D, see Chapter 10, “Shared Memory Communications,” on page 503.

Use the ECSALIMIT parameter on the GLOBALCONFIG statement to limit TCP/IP use of common storage. The POOLLIMIT parameter can be used to limit TCP/IP use of private storage pools.

Use ZERT to enable z/OS Encryption Readiness Technology (zERT). For more information, see “Monitoring cryptographic network protection: z/OS encryption readiness technology (zERT)” on page 162.

IPCONFIG

Use IPCONFIG to configure various settings of the IP layer of TCP/IP. Use ARPTO to specify the ARP timeout value in seconds for LCS devices. For more information, see the ARPAGE description.

Use CLAWUSEDOUBLENOP on vendor devices that document the need for double NOPs on each CCW.

Use DATAGRAMFWD if this TCP/IP is to be a router and must forward datagrams to other routers. Use IGNOREREDIRECT when a dynamic routing program is used and ICMP redirect packets are to be ignored by the TCP/IP address space. MULTIPATH is used to inform TCP/IP how to distribute traffic across equal cost routes.

Use IPSECURITY to restrict this host to be a network firewall.

SOURCEVIPA enables interface fault tolerance for z/OS clients that establish outbound connections. When SOURCEVIPA is set, outbound datagrams use the corresponding virtual IP address (VIPA) in the HOME list instead of the physical interfaces IP address. SOURCEVIPA has no effect on RIP servers such as OMPROUTE.

TCPSTACKSOURCEVIPA allows z/OS clients to specify a sysplex-wide source IP address for TCP connections. When TCPSTACKSOURCEVIPA is set, outbound TCP datagrams use the IP address that is specified in the TCPSTACKSOURCEVIPA statement instead of static VIPA addresses or physical interface addresses.

Use SYSPLEXROUTING to communicate interface changes within a sysplex domain to the workload manager (WLM). DYNAMICXCF allows the cross communication facility within a sysplex to dynamically generate connections within a sysplex domain. If DYNAMICXCF is used with a dynamic routing program like OMPROUTE, the BSDROUTINGPARMS and the OMPROUTE configuration files must be updated with subnet mask and cost information. For more information about other configuration parameters that are required, see the usage notes related to the DYNAMICXCF parameter under the IPCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Use REASSEMBLYTIMEOUT to specify the TCP/IP reassemble timeout value in seconds, and the TTL specifies the TCP/IP time to live or hop count value.

Use PATHMTUDISCOVERY to indicate to TCP/IP that it is to dynamically discover the path MTU, which is the minimum of MTUs of each hop in the path.

Use STOPONCLAWERROR to indicate to the TCP/IP stack to stop channel programs (HALTIO and HALTSIO) when a device error is detected.

Use QDIOACCELERATOR to request accelerated packet forwarding for OSA-Express QDIO Ethernet and HiperSockets interfaces.

IPCONFIG6

Use IPCONFIG6 to update the IP layer of TCP/IP with information that pertains to IPv6.

Use DATAGRAMFWD to enable the transfer of data between networks.

Use DYNAMICXCF to enable Dynamic XCF support for IPv6.

SOMAXCONN

Use SOMAXCON to specify the maximum number of sockets queued on a listener.

SRCIP

Use the SRCIP - ENDSRCIP profile statement block to configure one of the following functions:

- Enable an application to use a designated IP address as its source address for outbound TCP connections, or to enable a TCP server application to bind to a specific IP address when it is establishing its listening socket.
- Indicate that the default source address selection algorithm prefers public or temporary IPv6 addresses for specific jobs.

For outbound TCP connections, when a source IP address was designated for a specified job name or destination address and the source IP address exists at the time the outbound TCP connection is initiated, this source IP address is used, overriding other source IP address selection methods as described in "Source IP address selection" on page 244. This source address selection occurs for applications that issue a connect() call and that did not previously bind the socket to an IP address, or for those applications that bind to the IPv4 INADDR_ANY address or to the IPv6 unspecified address (in6addr_any) before they issue the connect() call.

For TCP server applications, when the application issues a bind to INADDR_ANY or in6addr_any and a matching JOBNAME rule for SERVER or BOTH is specified, the designated IP address is used on the listening socket. This situation makes the server application bind specific, where client applications can connect to the server by using only the designated IP address. This capability can be useful when the applications do not provide a method for the user to specify a specific IP address for their listening sockets, or in situations when the server application creates listening sockets by using an ephemeral port that is assigned dynamically by TCP/IP. For scenarios when the application binds to specific, well-known ports, the BIND keyword on the PORT reservation statement in the TCP/IP profile can be used instead and has precedence over the SRCIP block specifications.

If you use distributed DVIPAs as a designated source within the SRCIP block, you might also be required to specify the EXPLICITBINDPORTRANGE parameter on the GLOBALCONFIG statement. For more information about the GLOBALCONFIG statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

Guidelines:

- Applications that bind to INADDR_ANY or in6addr_any that match on an SRCIP JOBNAME or DESTINATION statement do not have the designated IP address as their source address upon completion of the bind() call. The source address is not set to the designated address until completion of the subsequent connect() (client applications) or listen() (server applications) call. This situation is important to note for applications that issue a getsockname() call after a bind() call to retrieve the source IP address. This processing is different from the processing that occurs when a TCP server application is converted to being bind specific using the BIND keyword on the PORT statements in the TCP/IP profile. When you are using the BIND keyword on the PORT statement, the designated IP address is set upon completion of the bind() call, and some applications such as Db2 depend on this behavior.
- When you are using an SRCIP JOBNAME statement for an IPv6 server application, code an IPv6 address and not an IPv6 interface. Otherwise, the source address that is chosen for that IP interface might not be the best choice for the server application to be bound to. For information about the default source address selection algorithm, see z/OS Communications Server: IPv6 Network and Application Design Guide.

TCPCONFIG

Use the TCPCONFIG statement to configure various settings of the TCP protocol layer:

- Use the INTERVAL parameter if necessary to change the default keepalive value to a value other than 120 minutes. Use the KEEPALIVEPROBES parameter to specify the number of probes to send before a connection times out. Use the KEEPALIVEPROBEINTERVAL parameter to specify the amount of time between the sending of each probe.
- Use the FINWAIT2TIME parameter to specify a different timeout value for a TCP connection that is in the FINWAIT2 state.
- Use the TIMEWAITINTERVAL parameter to specify a different timeout value for a TCP connection that is in the TIMEWAIT state.

- Use the SENDGARBAGE parameter to cause the keepalive packet to contain 1 byte of random data and an incorrect sequence number. The random data and incorrect sequence number assure that the remote TCP does not accept the data.
- Use the TCPTIMESTAMP parameter to choose whether to participate in time stamp negotiation.
- Use the MAXIMUMRETRANSMITTIME parameter to limit the length of time before a connection times out.
- Use the RETRANSMITATTEMPTS parameter to indicate the number of packets to retransmit before a connection times out.
- Use the CONNECTTIMEOUT parameter to limit the amount of time before the initial connection times out.
- Use the CONNECTINITINTERVAL parameter to specify the initial retransmit interval for a connect call.
- Use the QUEUEDRTT parameter to specify the round trip time that triggers outbound serialization logic.
- Use the FRRTHRESHOLD parameter to specify the number of duplicates that are needed to trigger Fast Retransmit, Fast Recovery logic.
- Use the DELAYACKS parameter to alter the behavior of acknowledgments and delay their transmission.
- Use the NONAGLE parameter to override use of the Nagle algorithm. The Nagle algorithm is used to delay small packets from being sent.
- If you specify the RESTRICTLOWPORTS parameter, only applications that meet at least one of the following criteria are allowed to bind to low ports (1–1023):
  - The port is reserved for the application by the PORT or PORTRANGE statement.
  - The application runs with APF authorization.
  - The application runs with effective POSIX UID zero.
- If you want to control TCP buffering to limit storage usage or to manage large bandwidth devices, use the TCPSENDBFRSIZE, TCPRCVBUFRSIZE, TCPMAXSENDBUFRSIZE, and TCPMAXRCVBUFRSIZE parameters.
- Use the TTLS parameter to configure the TCP/IP stack for AT-TLS support.
- Use the EPHEMERALPORTS parameter to limit the ephemeral port range that the TCP/IP stack uses to assign a port to a socket. The EPHEMERALPORTS port range is used in the following situations when EXPLICITBINDPORTRANGE, SYSPLEXPORTS, or FTP PASSIVEDATAPORTS processing cannot determine the port number to assign:
  - An application issues an explicit bind() call for port 0
  - An application bypasses issuing an explicit bind() call and issues a connect() call
- The SELECTIVEACK parameter causes the TCP/IP stack to generate selective acknowledgments as defined in RFC 2018 and to use incoming selective acknowledgments to improve TCP retransmission processing as defined in RFC 3517. A TCP connection can experience poor performance when multiple packets are lost from one window of data. With the limited information available from cumulative acknowledgments, a TCP sender can learn about only a single lost packet per round-trip time. A Selective Acknowledgment (SACK) mechanism with a selective repeat retransmission policy can help to overcome these limitations. The receiving TCP sends back SACK packets to the sender to inform the sender of data that was received. The sending TCP can then retransmit only the missing data segments.

UDPCONFIG

Use UDPCONFIG to configure various settings of the UDP protocol layer. NOUDPCHKSUM can be used to eliminate check summing overhead for IPv4 UDP packets. This option is ignored for UDP datagrams that are flowing over an IPv6 network, as UDP Checksum is a required function on an IPv6 network.

If RESTRICTLOWPORTS is specified, only applications that meet at least one of the following criteria are allowed to bind to low ports (1–1023):

- The port is reserved for the application by the PORT or PORTRANGE statement.
- The application runs with APF authorization.
- The application runs with effective POSIX UID zero.

If an installation wants to control UDP buffering (to limit storage usage or to manage large bandwidth devices), use the UDPSENDBFRSIZE and UDPRCVBUFRSIZE parameters. UDPQUEUELIMIT can be used to set a queue limit for UDP. UDPQUEUELIMIT is useful for installations that want to limit the size of the queue of UDP datagrams that an application can have waiting before the TCP/IP address space starts discarding them.

Use EPHEMERALPORTS to limit the ephemeral port range that the TCP/IP stack uses to assign a port to a socket in the following situations:

- An application issues an explicit bind() call for port 0
- An application bypasses issuing an explicit bind() call

Source IP address selection

TCP/IP determines the source IP address for a TCP outbound connection, or for a UDP or RAW outbound packet, by using the following sequence, which is listed in descending order of priority.

1. Sendmsg( ) using the IPV6_PKTINFO ancillary option specifying a nonzero source address (RAW and UDP sockets only)
2. Setsockopt( ) IPV6_PKTINFO option specifying a nonzero source address (RAW and UDP sockets only)
3. Explicit bind to a specific local IP address
4. bind2addrsel socket function (AF_INET6 sockets only)
5. PORT profile statement with the BIND parameter
6. SRCIP profile statement (TCP connections only)
7. TCPSTACKSOURCEVIPA parameter on the IPCONFIG or IPCONFIG6 profile statement (TCP connections only)
8. SOURCEVIPA: Static VIPA address from the HOME list or from the SOURCEVIPAINTERFACE parameter
9. Local IP address of the interface over which the packet is sent

For a TCP connection, the source address is selected for the initial outbound packet, and the same source IP address is used for the life of the connection. For the UDP and RAW protocols, a source IP address selection is made for each outbound packet. A more detailed description of this sequence follows.

How TCP/IP selects a source IP address

TCP/IP uses the following sequence to select the source IP address for an outbound packet. For detailed information about the TCP/IP profile (PROFILE.TCPIP) and configuration statements, see z/OS Communications Server: IP Configuration Reference.

1. Sendmsg that specifies the source address in the ancillary IPV6_PKTINFO data

   If this is a UDP or RAW socket, and IPV6_PKTINFO ancillary data is specified on sendmsg() with a nonzero source IP address, this address is used as the source IP address.

2. Setsockopt IPV6_PKTINFO

   If this is a UDP or RAW socket, and the IPV6_PKTINFO socket option is set and it contains a nonzero source IP address, this address is used as the source IP address.

3. Explicit bind to a specific local address

   If the socket is already bound to a specific local IP address other than INADDR_ANY or in6addr_any, TCP/IP uses this specific local IP address.

4. bind2addrsel socket function (IPv6 only)

The bind2addrsel socket function, which is available only to AF_INET6 sockets, binds a socket to a stack-selected local address and port that is appropriate to communicate with a given destination address.

5. PORT profile statement with the BIND parameter

   If the socket is bound to a source port and to the INADDR_ANY or in6addr_any IP address, and there is a corresponding PORT profile statement with the BIND parameter specified, TCP/IP uses the address specified by the BIND parameter.

6. SRCIP profile statement

   If this is a TCP socket and either the socket is not yet bound or the socket is bound to the INADDR_ANY or in6addr_any IP address, TCP/IP checks the job name and destination IP address against the SRCIP entries in the following order:

   a. JOBNAME entries, other than JOBNAME *

   b. DESTINATION entries

   c. JOBNAME * entries

   If a match is found, TCP/IP uses the designated source in the most specific matching entry to provide the source IP address to be used.

   Restriction: JOBNAME entries that specify a source of PUBLICADDRS or TEMPADDRS apply to only IPv6 source IP address selection. The JOBNAME entries do not influence the setting of the source IP address during this step of the source IP address selection sequence. For information about how PUBLICADDRS or TEMPADDRS entries influence the selection of the source IP address and the default source address selection algorithm, see z/OS Communications Server: IPv6 Network and Application Design Guide.

7. TCPSTACKSOURCEVIPA parameter on the IPCONFIG or IPCONFIG6 profile statement

   All of the following conditions must be met:

   - This is a TCP socket.
   - SOURCEVIPA is enabled on IPCONFIG or IPCONFIG6.
   - SOURCEVIPA is not disabled for the socket.
   - The application has not issued a specific bind for this socket, even to the INADDR_ANY or in6addr_any IP address.
   - The address specified on the TCPSTACKSOURCEVIPA parameter is a static VIPA or active dynamic VIPA.

     For a static VIPA, it is defined on a DEVICE/LINK/HOME statement and not the INTERFACE statement.

   If these conditions are met and this is an IPv4 packet, TCP/IP uses the address specified on the TCPSTACKSOURCEVIPA parameter.

   If these conditions are met and this is an IPv6 packet, TCP/IP uses the default source address selection algorithm to select one of the addresses configured for the VIPA interface referenced by the TCPSTACKSOURCEVIPA parameter. For information about the default source address selection algorithm, see z/OS Communications Server: IPv6 Network and Application Design Guide.

   Guideline: Because the SRCIP profile statement provides all of the functionality of the TCPSTACKSOURCEVIPA parameter and additional granularity, consider using the SRCIP statement instead of specifying the TCPSTACKSOURCEVIPA parameter. Specifying JOBNAME * in a SRCIP profile statement provides the same result as specifying the TCPSTACKSOURCEVIPA parameter for implicit bind scenarios, and also applies to applications that issue a bind to the INADDR_ANY or in6addr_any IP address.

8. SOURCEVIPA: Static VIPA address from the HOME list (IPv4 interface defined with the LINK statement) or from the SOURCEVIPAINTERFACE parameter (IPv6 or IPv4 interface defined with the INTERFACE statement)

All of the following conditions must be met:

- SOURCEVIPA is enabled on IPCONFIG or IPCONFIG6.
- SOURCEVIPA is not disabled for the socket.
- Either the socket is not yet bound, or the socket is bound to the INADDR_ANY or in6addr_any IP address.

If these conditions are met, TCP/IP determines the interface over which the initial packet will be sent.

- For an IPv4 packet sent over an interface that is defined with the LINK statement, TCP/IP takes the following actions:
  a. Locates that interface in the HOME list.
  b. Starting with that interface, searches backward in the HOME list for the nearest static VIPA.
  c. If a static VIPA is found in the HOME list, TCP/IP uses that static VIPA found as the source IP address.
- For an IPv4 packet sent over an interface that is defined with the INTERFACE statement, TCP/IP takes the following actions:
  a. Determines whether a SOURCEVIPAINTERFACE parameter was specified for the selected interface.
  b. If a SOURCEVIPAINTERFACE parameter was specified, TCP/IP uses the address of the VIPA interface that is referenced by the SOURCEVIPAINTERFACE parameter.
- For an IPv6 packet, TCP/IP takes the following actions:
  a. Determines whether a SOURCEVIPAINTERFACE parameter was specified for the selected interface.
  b. If a SOURCEVIPAINTERFACE parameter was specified, TCP/IP uses the default source address selection algorithm to select one of the addresses that are configured for the VIPA interface that is referenced by the SOURCEVIPAINTERFACE parameter. For information about the default source address selection algorithm, see z/OS Communications Server: IPv6 Network and Application Design Guide.

9. Local IP address of the interface over which the packet is sent

   For an IPv4 packet, TCP/IP uses the local IP address of the interface over which the initial packet is sent.

   For an IPv6 packet, TCP/IP uses the default source address selection algorithm to select one of the addresses configured for the interface over which the initial packet is sent. For information about the default source address selection algorithm, see z/OS Communications Server: IPv6 Network and Application Design Guide.

Ephemeral port selection

The sequence that the stack uses to determine the ephemeral port for an outbound connection or packet depends on the protocol that is being used. For an outbound UDP packet, the following hierarchy is used:

1. The port number is specified on an explicit bind.
2. The port number is assigned from the pool of ports that the EPHEMERALPORTS parameter on the UDPCONFIG statement defines.

   The default ephemeral port numbers are in the range 1024 – 65535.

For an outbound TCP connection, ephemeral port selection can be affected by source IP address selection during bind processing. For more information about source IP address selection, see “Source IP address selection” on page 244.

After a source IP address is selected, the ephemeral port is selected by using the following hierarchy:

1. The port number is specified on an explicit bind.

2. The port number is assigned from the pool of ports that the EXPLICITBINDPORTRANGE parameter on the GLOBALCONFIG statement defines.

   Rule: EXPLICITBINDPORTRANGE ports are used when an application issues an explicit bind() to port 0 and either the IPv4 address INADDR_ANY or the IPv6 unspecified address (in6addr_any), before the application issues a connect() request. EXPLICITBINDPORTRANGE ports are unique throughout the sysplex.
3. The port number is assigned from the PASSIVEDATAPORTS range. The PASSIVEDATAPORTS range must also be reserved on a PORTRANGE statement with the AUTHPORT parameter.

   Rule: The PASSIVEDATAPORTS range is used by the FTP server only, and specifies the range of ports that are used as listening data socket ports.
4. The port number is assigned by SYSPLEXPORTS processing when a client application binds to a distributed DVIPA with SYSPLEXPORTS specified and port 0.
5. The port number is assigned from the pool of ports that the EPHEMERALPORTS parameter on the TCPCONFIG statement defines.

   The default ephemeral port numbers are in the range 1024 – 65535.

For more information about ephemeral port selection and the interaction between the different methods of assigning ports, see “Port selection interactions” on page 247.

Port selection interactions

You can assign or reserve ports by using several methods, and the various methods interact in the following ways:

- EPHEMERALPORTS parameter on the TCPCONFIG and UDPCONFIG statements
  - For TCP sockets, ephemeral ports are port numbers that TCP/IP temporarily assigns to a socket in the following situations:
    - An application explicitly binds to port 0.
    - An implicit bind() call is processed as an application issues a connect() call, and the port number is not known.

    Ephemeral ports can also be used by servers in some cases, such as the FTP server in passive mode.
  - For UDP sockets, ephemeral ports are port numbers that TCP/IP temporarily assigns to a socket in the following situations:
    - An application explicitly binds to port 0.
    - An implicit bind() call is processed as the application sends data, and the port number is not known.
  - You can use the EPHEMERALPORTS parameter to control the range of ephemeral port numbers that the stack assigns.
  - Separate definitions for the TCPCONFIG and UDPCONFIG statements enable different ranges per protocol.
  - The default ephemeral port numbers for each protocol are port numbers in the range 1024 – 65535.
  - Generally, other methods of assigning ports take precedence over assigning ephemeral ports from the EPHEMERALPORTS range. The ports available for ephemeral port assignment are those ports in the EPHEMERALPORTS range that are left after other methods of assigning ports are applied.
- PORT and PORTRANGE statement
  - Ports can be reserved for specific uses with the PORT and PORTRANGE statements.
  - Ports can be reserved for a protocol by the PORT or PORTRANGE statement and are not available as ephemeral ports for that protocol. If the ports are also within the defined EPHEMERALPORTS range for that protocol, the stack cannot assign them as ephemeral ports.
- EXPLICITBINDPORTRANGE parameter on the GLOBALCONFIG statement

  - The EXPLICITBINDPORTRANGE parameter designates a range of unique ports across a sysplex. A participating stack can allocate a port from this range when an application binds explicitly to port 0 and either of the following addresses:
    - IPv4 address INADDR_ANY
    - IPv6 unspecified address (in6addr_any)
  - The most recently processed profile or VARY TCPIP,,OBEYFILE command that specifies the EXPLICITBINDPORTRANGE parameter defines the range of ports for the sysplex.
  - An explicitly bound port that is within the EXPLICITBINDPORTRANGE pool is not available for assignment by the stack as an EXPLICITBINDPORTRANGE port.
  - The EXPLICITBINDPORTRANGE pool is not required to be within the EPHEMERALPORTS range.
  - The EXPLICITBINDPORTRANGE pool takes precedence over the EPHEMERALPORTS range. Ports in the EXPLICITBINDPORTRANGE pool that are also within the defined EPHEMERALPORTS range cannot be assigned by the stack as ephemeral ports. Ports in the EXPLICITBINDPORTRANGE pool are available only for EXPLICBINDPORTRANGE use.
- SYSPLEXPORTS parameter on the VIPADISTRIBUTE statement
  - EPHEMERALPORTS processing controls SYSPLEXPORTS allocation.
    - If a bind is made to a distributed DVIPA with SYSPLEXPORTS and port 0 specified, the coupling facility assigns a block of sysplex ports from the EPHEMERALPORTS range to the stack for use with the DVIPA. The stack chooses a port from this block.
    - If the bind specifies a specific port from the EPHEMERALPORTS range, the coupling facility attempts to assign that port as a sysplex port for that DVIPA.
  - TCP/IP communicates the EPHEMERALPORTS range to the coupling facility to ensure that sysplex port assignments are in the correct range.
- FTP PASSIVEDATAPORTS statement and the AUTHPORT parameter on the PORTRANGE statement
  - The PASSIVEDATAPORTS statement in the FTP server configuration file assigns a range of port numbers for use by the FTP server. The PASSIVEDATAPORTS range must also be reserved on a PORTRANGE statement with the AUTHPORT parameter.
  - Ports in this range are available for assignment only to the FTP server.
  - Ports in this range are not required to be within the EPHEMERALPORTS range.
  - Ports in this range that are also within the EPHEMERALPORTS range are not available for use by the stack when assigning an ephemeral port.
- INADDRANYPORT and INADDRANYCOUNT parameters on the NETWORK statement in BPXPRMxx
  - The range that is specified by the INADDRANYPORT and INADDRANYCOUNT parameters must be restricted by the PORT or PORTRANGE statement to the job name OMVS. These ports are not assigned by the stack unless the user has the job name OMVS.
  - Ports in this range are not required to be within the EPHEMERALPORTS range.
  - Ports in this range that are within the defined EPHEMERALPORTS range are not available for use by the stack when assigning an ephemeral port.

Because ports that are reserved by using some of these methods cannot be used by the stack as ephemeral ports, the actual number of available ephemeral ports might not equal the number of ports that are specified in the EPHEMERALPORTS range. You can use the Netstat STATS/-S command to display the actual number of ephemeral ports that are available for assignment by the stack. The `CONFIGURED EPHEMERAL PORTS` field of the Netstat STATS/-S report accounts for ports that are in the EPHEMERALPORTS range that are unavailable because they are reserved by some other method. This field reflects the actual number of ports that are available for assignment by the stack.

For example, consider the following portion of a TCP/IP profile:

```
GLOBALCONFIG EXPLICITBINDPORTRANGE 30200 100
;
TCPCONFIG EPHEMERALPORTS 30000 39999
```

```
;
UDPCONFIG EPHEMERALPORTS 30000 39999
;
PORT      30005       TCP TCPAPPL2
PORT      30015       TCP TCPAPPL3
;
PORTRANGE 30025 5     TCP TCPAPPL6
PORTRANGE 30100 50    TCP AUTHPORT   ; FTP passive data ports
PORTRANGE 30500 300   TCP OMVS       ; INADDRANYPORT/INADDRANYCOUNT ports
;
PORT      30010       UDP UDPAPPL2
PORT      30015       UDP UDPAPPL3
;
PORTRANGE 30030 5     UDP UDPAPPL6
PORTRANGE 30500 300   UDP OMVS       ; INADDRANYPORT/INADDRANYCOUNT ports
```

This configuration results in the following values for the CONFIGURED EPHEMERAL PORTS field of the Netstat STATS/-S report:

```
d tcpip,tcpcs1,netstat,stats
EZD0101I NETSTAT CS V2R1 TCPCS1
.
.
.
TCP STATISTICS
.
.
.
  CONFIGURED EPHEMERAL PORTS          = 9543
  EPHEMERAL PORTS IN USE              = 0
  EPHEMERAL PORTS MAX USAGE           = 0
  EPHEMERAL PORTS EXHAUSTED           = 0
UDP STATISTICS
.
.
.
  CONFIGURED EPHEMERAL PORTS          = 9693
  EPHEMERAL PORTS IN USE              = 0
  EPHEMERAL PORTS MAX USAGE           = 0
  EPHEMERAL PORTS EXHAUSTED           = 0
END OF THE REPORT
```

For TCP, the total number of configured ephemeral ports is 10000 (ports 30000 – 39999), and this number is reduced by the following amounts:

- 100 because of the GLOBALCONFIG EXPLICITBINDPORTRANGE setting (ports 30200 – 30299)
- 2 because of the two PORT statements with ports that are within the TCP ephemeral port range (ports 30005 and 30015)
- 355 because of the three PORTRANGE statements with ports that are within the TCP ephemeral port range (ports 30025 – 30029, 30100 – 30149, and 30500 – 30799)

Subtracting these ports results in the CONFIGURED EPHEMERAL PORTS value of 9543. The CONFIGURED EPHEMERAL PORTS value is not affected by any PORT and PORTRANGE statements that include ports outside of the EPHEMERALPORTS range.

For UDP, the total number of configured ephemeral ports is 10000 (ports 30000 – 39999), and this number is reduced by the following amounts:

- 2 because of the two PORT statements with ports that are within the UDP ephemeral port range (ports 30010 and 30015)
- 305 because of the two PORTRANGE statements with ports that are within the UDP ephemeral port range (ports 30030 – 30034 and 30500 – 30799)

Subtracting these ports results in the CONFIGURED EPHEMERAL PORTS value of 9693. The CONFIGURED EPHEMERAL PORTS value is not affected by any PORT and PORTRANGE statements that include ports outside of the EPHEMERALPORTS range.

Setting up physical characteristics in PROFILE.TCPIP

Figure 45 on page 254 shows a portion of the sample configuration file for the TCP/IP address space, PROFILE.TCPIP. Figure 45 on page 254 includes the portion of the sample that shows how to set up physical characteristics. This sample can be copied from SEZAINST(SAMPPROF). Following Figure 45 on page 254, several of the statements that are used to set up physical characteristics in PROFILE.TCPIP are described.

```
; ====================================================================
; Network interface definitions
; ====================================================================
;
; DEVICE: Defines name (and sometimes device number) for various types
;   of network interfaces for IPv4 only
; LINK: Defines a network interface to be associated with a particular
;   device.  For IPv4 only.
; INTERFACE: Defines an IPv6 interface, or an IPv4 OSA-Express QDIO
;   ethernet, HiperSockets, or static virtual interface.
; VIPADYNAMIC: Defines IPv4 and IPv6 dynamic virtual interfaces
;
; --------------------------------------------------------------------
; DEVICE and LINK statements
; (Note: Also see INTERFACE statements for IPv4)
; --------------------------------------------------------------------
;
; DEVICE and LINK for CTC devices
;
;DEVICE  CTC1    CTC  D00  AUTORESTART
;LINK    CTCD00  CTC  0  CTC1
;
; DEVICE and LINK for LAN Channel Station and OSA devices:
;   DEVICE: Defines name and hexadecimal device number for an IBM 8232
;     LAN channel station (LCS) device, and IBM 3172 Interconnect
;     Controller, an IBM 2216 Multiaccess Connector Model 400,
;     an Ethernet.
;   LINK: Defines a network interface link associated with an LCS
;     device; may be for Ethernet Network.
;
;   Example: LCS1 is a 3172 model 1 with an Ethernet
;   adapter
;DEVICE  LCS1   LCS  BA0  AUTORESTART
;LINK ETH1 ETHERNET 1 LCS1
;
; DEVICE and LINK for MPCIPA QDIO Devices:
;
;   Example: MPCIPA1 is an OSA-Express QDIO Ethernet feature
;
;DEVICE  MPCIPA1     MPCIPA  NONROUTER AUTORESTART
;LINK    MPCIPALINK1 IPAQENET  MPCIPA1
;
;   Example: MPCIPA2 is an OSA-Express QDIO Ethernet feature
;   configured as the PRIMARY router
;
;DEVICE  MPCIPA2     MPCIPA  PRIROUTER AUTORESTART
;LINK    MPCIPALINK2 IPAQENET  MPCIPA2
;
; DEVICE and LINK for HiperSockets CHPID FE
;
;DEVICE  IUTIQDFE      MPCIPA    AUTORESTART
;LINK    HIPERSOCKFE4  IPAQIDIO  IUTIQDFE
;
; DEVICE and LINK for MPCPTP devices:
;
;DEVICE  MPCPTP1    MPCPTP   AUTORESTART
;LINK    MPCPTPLINK MPCPTP   MPCPTP1
;


;
;
; DEVICE and LINK for static virtual (VIPA) interfaces:
;
;DEVICE  VDEV1   VIRTUAL  0
;LINK    VLINK1  VIRTUAL  0  VDEV1
;

; --------------------------------------------------------------------
; INTERFACE statements for IPv4
; --------------------------------------------------------------------
```

```
;  INTERFACE VIPAINT1                  ; IPv4 static virtual (VIPA)
;    DEFINE VIRTUAL
;    IPADDR 9.67.116.100
;  -------------------------------------
;  INTERFACE OSAQDIO24                 ; IPv4 OSA-Express QDIO ethernet
;    DEFINE IPAQENET
;    PORTNAME OSAQDIO2
;    INBPERF DYNAMIC
;    VMAC
;    SOURCEVIPAINT VLINK1
;    IPADDR 9.67.113.84/24             ; address and subnet mask
;  -------------------------------------
;  INTERFACE HIPERSOCK1                ; IPv4 HiperSockets
;    DEFINE IPAQIDIO
;    CHPID  FC
;    SOURCEVIPAINT VIPAINT1
;    IPADDR 9.67.122.10/24             ; address and subnet mask
;
; ------------------------------------------------------------------------
; INTERFACE statements for IPv6
; ------------------------------------------------------------------------
;  IPv6 Virtual interface definitions
;  ----------------------------------
;      IPADDR keyword is required for Virtual interfaces
;      Multiple IP addresses can be defined to one interface
;      The prefixes of the IPv6 VIPA addresses should be
;        different than the prefixes used for addresses
;        configured or autoconfigured for real interfaces.
;  ------------------------------------------------------------------------
;  INTERFACE VIPAV6 DEFINE
;    VIRTUAL6
;    IPADDR 50C9:C2D4:0:A:9:67:115:66 ; (Global Address)
; ------------------------------------------------------------------------
;  OSA-Express QDIO interface definitions
;  --------------------------------------
;   To use autoconfiguration for IPv6 interfaces, the IPADDR
;     parameter cannot be specified.
;   To manually define address(es), use the IPADDR keyword.
;   To assign a source VIPA address for an interface, use SOURCEVIPAINT
;   For OSA-Express QDIO ethernet features, you can define both
;     the IPv4 and IPv6 interfaces with INTERFACE statements, but you
;     must define a datapath device for each interface.
;   To configure a virtual MAC address for an OSA-Express QDIO ethernet
;     feature, specify the VMAC parameter on the INTERFACE
;     statement.
;  -------------------------------------
;  INTERFACE OSAQDIO26                 ; IPv6 OSA-Express QDIO ethernet
;    DEFINE IPAQENET6
;    PORTNAME OSAQDIO2
;    INBPERF DYNAMIC
;    VMAC
;    SOURCEVIPAINT VIPAV6
;    IPADDR 50C9:C2D4:0:1:9:67:115:66 ; (Global Address)
;
; ------------------------------------------------------------------------
;  MPC Point-to-point interface definitions
;  ----------------------------------------
;  MPCPTP6 interfaces require manual IPv6 address configuration. If
;  you don't code an IPADDR on the MPCPTP6 interface, you'll get only
;  the link-local address.  Here, the specified interface ID (INTFID)
;  will be appended to the global prefix, to form the global
;  address.
;  -------------------------------------
;  INTERFACE MPC1IPV6          ; MPCPTP6 (IPv6 over MPC point to point)
;    DEFINE MPCPTP6
;    TRLENAME TRLE1A
;    INTFID 0012:3456:789A:BCDE
;    IPADDR 2001:0DB8:0:0/64   ; (Global Prefix)
; ------------------------------------------------------------------------
;  HiperSockets IPv6
;  -----------------
;  INTERFACE HIPERSOCKFE6     ; IPAQIDIO6 (HiperSockets IPv6)
;    DEFINE IPAQIDIO6
;    CHPID  FE
;    INTFID 0000:2000:2000:2000
;    IPADDR 50C9:C2D4:0:1:9:67:118:80 ; (Global Address)
; ------------------------------------------------------------------------
;  To define other IPv6 Loopback addresses:
;  -----------------------------------------
;   INTERFACE LOOPBACK6 ADDADDR ::0014:0
;
; ------------------------------------------------------------------------
```

```
; VIPADYNAMIC statement for dynamic virtual (DVIPA) interfaces
; -------------------------------------------------------------------
;
; The VIPADYNAMIC statement provides the following functions:
;   - Define dynamic virtual (DVIPA) interfaces on a stack.
;   - Define sysplex distribution of TCP connections using DVIPAs.
;   - Define a stack as a backup for DVIPAs.
;   - Define the IP address range for creating DVIPAs by use of IOCTL
;     or Bind requests.
;   - Define routes over interfaces other than dynamic XCF, to be
;     used for the forwarding of DVIPA packets by sysplex
;     distributor.
;
; VIPADYNAMIC
;
;   Define two IPV4 dynamic VIPAs on this stack:
;   VIPADEFINE 255.255.255.192 201.2.10.11 201.2.10.12
;
;   Define two IPv6 dynamic VIPAs on this stack:
;   VIPADEFINE DVIPA1 1::1
;   VIPADEFINE DVIPA2 2::2
;
;   Define this stack as backup for these dynamic VIPAs and,
;   define the DVIPAs on this stack if they are not
;   already active elsewhere in the sysplex:
;   VIPABACKUP  200 MOVEABLE IMMEDIATE 255.255.255.192 9.67.240.02
;   VIPABACKUP  MOVEABLE IMMEDIATE V6DVIPA1 2000::9:67:240:2
;
;   Define Sysplex Distributor statements for these dynamic VIPAs:
;   VIPADISTRIBUTE SYSPLEXPORTS 201.2.10.11 PORT 4011 DESTIP ALL
;   VIPADISTRIBUTE TIMEDAFF 200 201.2.10.12 PORT 4012 DESTIP ALL
;   VIPADISTRIBUTE DISTMETHOD ROUNDROBIN DVIPA1       DESTIP ALL
;   VIPADISTRIBUTE SYSPLEXPORTS DVIPA2      PORT 4013 DESTIP ALL
;   VIPADISTRIBUTE
;     DISTMETHOD SERVERWLM  9.67.240.02 PORT 10000  DESTIP ALL
;   VIPADISTRIBUTE
;     DISTMETHOD SERVERWLM PROCXCOST ZIIP 2 ZAAP 2 ILWEIGHTING 2
;     OPTLOCAL 1  SYSPLEXPORTS
;     V6DVIPA1 PORT 10000  DESTIP ALL
;
;   Define this stack as backup for dynamic VIPAs on
;   other TCP/IP stacks:
;   VIPABACKUP 100             201.2.10.13 201.2.10.14
;   VIPABACKUP  80             201.2.10.21 201.2.10.22
;   VIPABACKUP  60             201.2.10.31 201.2.10.33
;   VIPABACKUP  40             201.2.10.32 201.2.10.34
;
;   Define two dynamic VIPA ranges on this stack:
;   VIPARANGE DEFINE 255.255.255.192 201.2.10.192
;   VIPARANGE V6RANGE1 3::1/100
;
;   A VIPARANGE statement with the optional keyword SAF followed by a
;   1-8 character name can be used to control the creation of DVIPA
;   addresses in the specified range with a security product such as
;   RACF. The full resource name for the security product authorization
;   check is constructed as follows:
;     EZB.BINDDVIPARANGE.sysname.tcpname.resname (for bind socket call)
;     EZB.MODDVIPA.sysname.tcpname.resname (for SIOCSVIPA ioctl,
;       SIOCSVIPA6 ioctl, or MODDVIPA utility)
;     where:
;       EZB.MODDVIPA and EZB.BINDDVIPARANGE are constant
;       sysname is the MVS system name (substitute your sysname)
;       tcpname is the TCPIP jobname (substitute your jobname)
;       resname is the 1-8 character name following the SAF keyword
;
;   VIPARANGE DEFINE 255.255.255.0 201.3.10.1 SAF APPL1
;   VIPARANGE V6RANGE1 6::1/64 SAF APPL2
;
;   Define alternative routes to dynamic XCF routes, for forwarding of
;   DVIPA packets by sysplex distributor
;   VIPAROUTE DEFINE 201.1.10.20     9.67.213.81
;   VIPAROUTE DEFINE 2001::251:00002 50C9:C2D4:0:1:9:67:215:66
; ENDVIPADYNAMIC
;
; -------------------------------------------------------------------
; Other interface statements
; -------------------------------------------------------------------
;
; TRANSLATE: Indicates a relationship between an internet address and
;   the network address on a specified link.  Only applicable for IPv4
;   interfaces defined by DEVICE and LINK profile statements.
;
```

```
; TRANSLATE
;   9.67.43.110   ETHERNet    FF0000006702   ETH1
;

;Destination               First Hop    Link Name Packet Size
; ======================================================================
; HOME addresses
; ======================================================================
;
; HOME: Provides the list of home IP addresses and associated link
;       names for IPv4 interfaces defined by DEVICE and LINK profile
;       statements.
;
;   - The LOOPBACK statement of 14.0.0.0 should only be used if the
;     installation has applications that require this old loopback
;     address.  The current stack uses 127.0.0.1 as the loopback
;     address.
;
; HOME
;  14.0.0.0       LOOPBACK
;  193.5.2.1      ETH1
;  9.67.113.80    CTCD00
;  9.67.113.81    MPCIPALINK1
;  9.67.113.82    MPCPTPLINK
;  9.67.113.83    MPCIPALINK2
;  9.67.116.86    VLINK1
;  9.67.100.80    LINK1
;  9.67.118.80    HIPERSOCKFE4
;
;
; PRIMARYINTERFACE: Specifies which link is designated as the default
;   local host for use by the GETHOSTID() function.  Only applicable
;   for IPv4 interfaces defined by a LINK or INTERFACE statement.
;
;   - If PRIMARYINTERFACE is not specified, then the first link in
;     the HOME statement is the primary interface, as usual.
; PRIMARYINTERFACE ETH1
;
;
; ======================================================================
; Routing configuration
; ======================================================================
; ----------------------------------------------------------------------
; Static routing
; ----------------------------------------------------------------------
;
; BEGINRoutes: Defines static routes to the main route table for IPv4
;                  and IPv6
;
;  - Use of BEGINROUTES with OMPROUTE routing daemon is not recommended.
;
;BEGINRoutes
;
; Direct Routes - Routes that are directly connected to my interfaces.
;
;      Destination Subnet Mask   First Hop    Link Name Packet Size
;
;ROUTE 193.5.2.0/24              =            ETH1      MTU 1500
;
;      Destination Subnet Mask   First Hop    Interface   Packet Size
;
;ROUTE FE80::1:2:3:4/128         =            OSAQDIO26     MTU 2000
;ROUTE 2001::0DB8:1/128          =            OSAQDIO26     MTU 2000
;
;
; Indirect Routes - Routes that are reachable through routers on my
;                   network.
;
;      Destination Subnet Mask   First Hop    Link Name Packet Size
;
;ROUTE 10.5.6.4    HOST          193.5.2.10   ETH1      MTU 1500
;
;      Destination Subnet Mask   First Hop      Interface   Packet Size
;
;ROUTE FEC8::/64                 FE80::1:2:3:4  OSAQDIO26   MTU 2000
;
; Default Route - All packets to an unknown destination are routed
;                 through this route.
;
;      Destination               First Hop    Link Name Packet Size
;
```

```
;ROUTE DEFAULT                   193.5.2.99   ETH1     MTU DEFAULTSIZE

;
;      Destination Subnet Mask First Hop      Interface Packet Size
;
;ROUTE DEFAULT6                FE80::1:2:3:4  OSAQDIO26 MTU DEFAULTSIZE
;
;ENDRoutes
;
; ----------------------------------------------------------------------
; Dynamic routing
; ----------------------------------------------------------------------
;
; BSDROUTINGPARMS: Defines the characteristics of each link defined at
;   the host. Only applicable for IPv4 interfaces defined by the
;   DEVICE and LINK profile statements.
;
; If not supplied, characteristics will be supplied from:
; (1) Static routing definitions in BEGINROUTES
; (2) OMPROUTE configuration (if OMPROUTE is running)
; (3) Stack's interface layer based on hardware capabilites and
;     characteristics of devices and links.
;
;   - OMPROUTE does not require BSDROUTINGPARMS.  OMPROUTE will
;     override the parameters with the coded or defaulted values
;     from its configuration.
;
; BSDROUTINGPARMS TRUE
;  Link name      MTU    Cost metric  Subnet Mask    Dest address
;   ETH1         1500         0       255.255.255.0     0
;   VLINK1    DEFAULTSIZE     0       255.255.255.0     0
;   CTCD00       65527        0       255.255.255.0  9.67.113.90
; ENDBSDROUTINGPARMS
```

Figure 45. Example of physical characteristics in PROFILE.TCPIP

The following list describes several of the statements that are shown in Figure 45 on page 254, which are used to set up physical characteristics in PROFILE.TCPIP. For more information about any of these statements, or information about statements that are not described, see z/OS Communications Server: IP Configuration Reference. For information that is specific to IPv6 support, see z/OS Communications Server: IPv6 Network and Application Design Guide.

DEVICE and LINK

Use DEVICE and LINK statements to define each IPv4 network interface to the TCPIP address space. See z/OS Communications Server: IP Configuration Reference for more details about the various network interfaces supported by TCP/IP.

CTC

Use the CTC DEVICE and LINK statements to define connectivity to another z/OS using channel-to-channel.

LAN Channel Station (LCS) DEVICE and LINK

Use the LCS DEVICE and LINK statements to define connectivity to an Ethernet LAN. LCS devices can have more than one adapter. Therefore, you can have more than one LINK statement an LCS DEVICE statement.

In configurations where multiple LCS and/or MPCIPA links onto the same LAN are defined, if the interface targeted by the ARP Request is inactive, one of the other active interfaces on the LAN will automatically take over responsibility for answering ARPs on behalf of the inactive interface. In this way, fault tolerance is achievable on the LAN without requiring a dynamic routing protocol.

TCP/IP supports ARP for VIPAs. In a flat network (one in which traffic flows directly between two endpoints without an intermediate router) using static routing with multiple interfaces onto the same LAN, you can achieve fault tolerance by defining a VIPA in the same subnet as the physical interfaces on the LAN. If a static route specifies a VIPA as the next hop IP address, the host or router will send an ARP for the VIPA. TCP/IP will reply to the ARP with the MAC address of one of the active physical interfaces on that LAN.

MPCIPA

Use the MPCIPA DEVICE and LINK statements to define either of the following connectivity options:

- LAN connectivity with the OSA-Express feature using the Queued Direct I/O (QDIO) interface

  Tip: You can also use the INTERFACE statement to define an IPv4 interface for OSA-Express QDIO Ethernet, which combines the definitions of the DEVICE, LINK, and HOME statements into a single statement.
- HiperSockets connectivity

For the OSA-Express QDIO interface, the MPCIPA device name must be the PORT name of the TRLE definition of the QDIO interface as described in z/OS Communications Server: SNA Resource Definition Reference. Device specifications for the type of IP routing supported are also specified on the MPCIPA DEVICE statement. These are also described in z/OS Communications Server: SNA Resource Definition Reference.

In configurations where multiple LCS and/or MPCIPA links onto the same LAN are defined, if the interface targeted by the ARP Request is inactive, one of the other active interfaces on the LAN will automatically take over responsibility for answering ARPs on behalf of the inactive interface. In this way, fault tolerance is achievable on the LAN without requiring a dynamic routing protocol.

TCP/IP supports ARP for VIPAs. In a flat network (one in which traffic flows directly between two endpoints without an intermediate router) using static routing with multiple interfaces onto the same LAN, you can achieve fault tolerance by defining a VIPA in the same subnet as the physical interfaces on the LAN. If a static route specifies a VIPA as the next hop IP address, the host or router will send an ARP for the VIPA. TCP/IP will reply to the ARP with the MAC address of one of the active physical interfaces on that LAN.

Use the IPCONFIG DYNAMICXCF statement to cause TCP/IP to automatically define and activate HiperSockets connectivity between each pair of TCP/IP stacks that is in the same CPC in the same z/OS sysplex.

MPCPTP

MPCPTP can be used to define any of the following connections:

- A connection to another host over a series of CTCs (in this case, the device name must be the name of a VTAM TRLE).
- An XCF connection to another TCP/IP in the same z/OS sysplex. For an XCF connection, the device name must be the CP name or SSCP name of the target VTAM on the other side of the XCF connection, and the VTAM ISTLSXCF major node must be active to start the device.
- An IUTSAMEH connection (with no need for any I/O devices) to another TCP/IP on the same z/OS system or to VTAM for Enterprise Extender. For an IUTSAMEH connection, the device name must be the reserved name IUTSAMEH. VTAM automatically activates the IUTSAMEH TRLE.

Use the IPCONFIG DYNAMICXCF statement to cause TCP/IP to automatically define and activate XCF connectivity between each pair of TCP/IP stacks in the same sysplex and IUTSAMEH connectivity between multiple TCP/IP stacks on the same z/OS.

VIRTUAL

Use VIRTUAL DEVICE to define a static virtual IP address (VIPA) interface.

The static virtual device requires DEVICE and LINK statements to define a device that is always started, can never be stopped, can be known within the network, yet requires no physical adapters. It is very useful to define VIPAs so that if a physical adapter loses its connection to the network, application traffic using the failed physical adapter can be rerouted over another interface to the network. To the network, the VIPA address appears to be one hop away from the TCP/IP address spaces. The network sends and receives datagrams to and from the physical interfaces to get to the VIPA address. For more information about VIPA, see Chapter 7, “Virtual IP Addressing,” on page 361.

INTERFACE (IPv4)

Use the INTERFACE statement to define IPv4 OSA-Express QDIO interfaces, HiperSockets interfaces, or static VIPA interfaces (instead of using DEVICE and LINK statements).

IPAQENET
Use the IPAQENET INTERFACE statement to define IPv4 LAN connectivity through the OSA-Express feature using QDIO.

IPAQIDIO
Use the IPAQIDIO INTERFACE statement to define HiperSockets IPv4 connectivity.

VIRTUAL
Use the VIRTUAL INTERFACE statement to define an IPv4 static VIPA.

INTERFACE (IPv6)
Use INTERFACE statements to define each IPv6 network interface to the TCPIP address space. See z/OS Communications Server: IP Configuration Reference for more details about the various network interfaces supported by TCP/IP.

IPAQENET6
Use the IPAQENET6 INTERFACE statement to define IPv6 LAN connectivity through the OSA-Express feature using QDIO.

IPAQIDIO6
Use the IPAQIDIO6 INTERFACE statement to define HiperSockets IPv6 connectivity.

Use the IPCONFIG6 DYNAMICXCF statement to cause TCP/IP to automatically define and activate HiperSockets IPv6 connectivity between each pair of TCP/IP stacks that is in the same CPC in the same z/OS sysplex.

MPCPTP6
Use the MPCPTP6 INTERFACE statement to define any of the following IPv6 connections:

- A connection to another host using ESCON channel-to-channel adapters.
- An XCF connection to another TCP/IP in the same z/OS sysplex.
- An IUTSAMEH connection (with no need for any I/O devices) to another TCP/IP on the same z/OS system or to VTAM for Enterprise Extender.

Use the IPCONFIG6 DYNAMICXCF statement to cause TCP/IP to automatically define and activate XCF connectivity between each pair of TCP/IP stacks in the same sysplex and IUTSAMEH connectivity between multiple TCP/IP stacks on the same z/OS.

VIRTUAL6
Use the VIRTUAL6 INTERFACE statement to define IPv6 static VIPAs.

VIPADYNAMIC
Use the VIPADYNAMIC block statement to define dynamic virtual IP address (DVIPA) interfaces. For more information about DVIPA interfaces, see Chapter 7, “Virtual IP Addressing,” on page 361.

TRANSLATE
Use TRANSLATE to indicate which LINK has specified network addresses for use as a static ARP table. The first TRANSLATE statement in a configuration data set replaces the entire ARP cache. Subsequent TRANSLATE statements add to the table. If you are using OSPF routing (OMPROUTE), see Chapter 6, “Routing,” on page 279 for more information about requirements for the TRANSLATE statement.

HOME
HOME lists the IP addresses and their associated LINK adapter. The first HOME statement within a configuration data set replaces the existing HOME list. If subsequent HOME statements are found within a configuration data set, add entries to the list.

Guideline: The order of the HOME list is important if IPCONFIG SOURCEVIPA is specified. For more information, see “Source IP address selection” on page 244. For more information about the HOME statement and for precautions when either the VIPA address or a physical adapter used as a source for the VIPA has an IP address that is the network address, see z/OS Communications Server: IP Configuration Reference.

PRIMARYINTERFACE
Use PRIMARYINTERFACE to specify which interface should be designated as the default local host for use by the GETHOSTID() function. If PRIMARYINTERFACE is not used, the first IP address in the

HOME list becomes the default local host address. If there are no HOME statements defined in the profile, PRIMARYINTERFACE defaults to the first IPv4 INTERFACE statement listed in the profile.

BEGINROUTES
Use the BEGINROUTES statement to add static routes to the IP route table.

After an interface has a DEVICE, LINK, and HOME statement, or an INTERFACE statement, it can be started with the START statement or the VARY TCPIP,,START command.

Devices that support ARP offload

Certain devices provide an ARP offload function that offloads all ARP processing to the adapter. The function provided by the adapter impacts the ability of TCP/IP to display ARP cache information or ARP counter statistics for these devices.

Note: ARP processing is relevant only for IPv4 LAN interfaces.

The following devices provide an ARP offload function and provide ARP cache data or ARP counters to TCP/IP.

- MPCIPA (OSA-Express Gigabit Ethernet) with a minimum required microcode level of [MCL] 401
- All other OSA-Express features in QDIO mode

Note: If multiple TCP/IP instances are sharing the device, the ARP data will represent all TCP/IP instances using the device. This information is provided to TCP/IP every 30 seconds from the device.

The MPCIPA (OSA-Express Gigabit Ethernet) device, with a microcode level earlier than [MCL] 401, provides an ARP offload function and does not provide any ARP cache data or ARP counters to TCP/IP.

Note: For IPv6 LAN interfaces, TCP/IP performs all the neighbor discovery processing, maintains the neighbor cache, and provides the ability to display neighbor cache information.

Interface-layer fault-tolerance for local area networks (interface-takeover function)

The TCP/IP stack in the z/OS Communications Server provides transparent fault-tolerance for failed (or stopped) IPv4 devices or IPv6 interfaces, when the stack is configured with redundant connectivity onto a LAN. This support is provided by the z/OS Communications Server interface-takeover function, and applies to IPv4 MPCIPA and LCS device types and to the IPv6 IPAQENET6 interface type.

At device or interface startup time, TCP/IP dynamically learns of redundant connectivity onto the LAN, and uses this information to select suitable backups in the case of a future failure of the device or interface. This support makes use of ARP flows (for IPv4 devices) or neighbor discovery flows (for IPv6 interfaces), so upon failure (or stop) of a device or interface, TCP/IP immediately notifies stations on the LAN that the original IPv4 or IPv6 address is now reachable through the backup's link-layer (MAC) address. Users targeting the original IP address will see no outage due to the failure, and will be unaware that any failure occurred.

Because this support is built upon ARP or neighbor discovery flows, no dynamic routing protocol in the IP layer is required to achieve this fault tolerance. To enable this support, you need only to configure redundancy onto the LAN:

- You need redundant LAN adapters.
- For IPv4, you must configure and activate multiple LINKs onto the LAN.
- For IPv6, you need to configure and start multiple INTERFACEs onto the LAN.

Restriction: An IPv4 device cannot back up an IPv6 interface, and an IPv6 interface cannot back up an IPv4 device.

Rule: If static routing is used, there needs to be a static route to the LAN subnet over each interface onto the LAN. There also needs to be a default route and routes to destinations not directly attached to the LAN over each interface.

The interface-layer fault-tolerance feature can be used in conjunction with VIPA addresses, where applications can target the VIPA address, and any failure of the real LAN hardware is handled by the interface-takeover function. This differs from traditional VIPA usage, where dynamic routing protocols are required to route around real hardware failures.

IPv6 considerations: Stateless autoconfiguration and duplicate address detection

IPv6 provides the capability of autoconfiguring addresses for an interface by using information that is provided by IPv6 routers. Descriptions of this function can be found in RFC 2461 and RFC 2462. The term *autoconfigured IP address* is used here to mean an IP address that is created as a result of information that is received from a router advertisement. z/OS TCP/IP allows autoconfiguration if no IP addresses are defined on the profile INTERFACE statement using the IPADDR keyword. If the INTERFACE statement contains IPADDR definitions, this indicates that the installation is defining its own IP addresses and autoconfiguration is not wanted. The term *manually configured addresses* is used here to describe the addresses that are defined using the IPADDR keyword.

TCP creates an autoconfigured IP address for an interface if all three of the following conditions are met:

- The interface is active.
- A valid router advertisement containing prefix information with the autonomous flag on is received over the interface.
- No manually configured home addresses are defined for the interface at the time the router advertisement is received.

The IP address that is created is formed by appending the interface ID generated by the stack to the prefix supplied by the router advertisement. Autoconfigured IP addresses can be identified in the Netstat HOME/-h report by the `Autoconfigured` flag. For more information on the interface ID generated by the stack, see z/OS Communications Server: IPv6 Network and Application Design Guide.

An autoconfigured IP address exists until one of the following events occur:

- The valid lifetime specified by the most recent router advertisement expires. When the valid lifetime expires, the autoconfigured address is removed. Existing connections using this address are terminated when subsequent activity occurs on the connection. The router advertisement that contains the valid lifetime for the autoconfigured address can also specify a preferred lifetime. The preferred lifetime indicates that the IP address can be freely used. When the preferred lifetime expires, the autoconfigured address is considered deprecated. The deprecated state indicates that another IP address should be used if available and provides a transition period before the valid lifetime expires. A deprecated IP address can be identified in the Netstat HOME/-h report by the `deprecated` flag.
- The installation activates a profile that contains a manually configured IP address on the same interface as the autoconfigured IP address (that is, the INTERFACE statement contains the ADDADDR keyword). If this occurs, any autoconfigured IP addresses on that interface are deleted and existing connections using this address are terminated when subsequent activity occurs on the connection. The manually configured addresses are added and duplicate address detection for the newly added IP addresses initiated, if applicable.

Duplicate address detection is the process described in RFC 2462 which verifies that IPv6 home addresses are unique on the local link before assigning them to an interface. Duplicate address detection is performed on all IPv6 IPAQENET6 home addresses, whether they are manually configured or autogenerated, unless the INTERFACE statement specifies DUPADDRDET 0. Duplicate address detection is not done for LOOPBACK6 or VIRTUAL6 addresses. The duplicate address detection process sends a multicast neighbor solicitation and waits a period of time to see if another neighbor indicates that the address is in use. By default, only one neighbor solicitation is sent and the length of time waited is approximately one second. If no neighbor responds in that interval, the address is considered unique and the interface will start using it. The number of neighbor solicitations sent by duplicate address detection can be modified by the DUPADDRDET keyword on the INTERFACE statement. The duration of the wait interval (awaiting a response from a neighbor already using the address) can be modified by information obtained from routers on the attached network.

Duplicate address detection occurs when the interface is started. Unless the INTERFACE statement indicated duplicate address detection is to be bypassed, IPv6 manually configured addresses are unavailable until the interface is started and duplicate address detection is completed without finding another node on the local link with the same address. Prior to activation of the interface, manually configured addresses are shown in the Netstat HOME/-h report as unavailable with the reason `DUPLICATE ADDRESS DETECTION PENDING`. While the duplicated address detection is actively in progress for an address, the Netstat HOME/-h report shows the address as unavailable with the reason `DUPLICATE ADDRESS DETECTION IN PROGRESS`. If another neighbor indicates the address is in use during the duplicate address detection process, message EZZ9780I is issued and the address is not made available to the interface. If the address that was found to be in use is a manually configured address, the address appears in the Netstat HOME/-h report as unavailable with the reason `DUPLICATE ADDRESS DETECTED`.

A link-local address is required to activate a QDIO IPv6 interface and will be generated automatically by the stack. The link-local address is generated using the link-local prefix and the interface ID. If the link-local address generated from the interface ID is determined to be a duplicate, the interface is not activated if:

- Autoconfigured addresses are allowed.
- A manually configured home address specifying only the prefix was specified on the INTERFACE statement. In such a case, the interface ID needs to be used to form the complete link-local address, and because the interface ID has been found to be in use on the network, the formed link-local address cannot be used.

If duplicate address detection fails on the link-local address and only fully configured manual addresses were specified on the INTERFACE statement, up to two attempts are made to create a unique link local address using a randomly generated value instead of the interface ID. If duplicate address detection succeeds using the randomly generated link-local address, message EZZ9784I is issued indicating the generated address and the interface is activated.

Setting up reserved port number definitions in PROFILE.TCPIP

Figure 46 on page 261 shows a portion of the sample configuration file for the TCP/IP address space, PROFILE.TCPIP. This sample can be copied from SEZAINST(SAMPPROF). Figure 46 on page 261 includes the portion of the sample that shows how to set up reserved port number definitions. Descriptions for the statements follow Figure 46 on page 261.

```
; ====================================================================
; Application configuration
; ====================================================================
;
; AUTOLOG: Supplies TCPIP with the procedure names to start and the
; time value to wait at TCP start up for any of those procedures
; to terminate if they are active.
;
; AUTOLOG 5
;  FTPD JOBNAME FTPD1       ; FTP Server
;  LPSERVE                  ; LPD Server
;  OMPROUTE                 ; OMPROUTE Server
;  OSNMPD                   ; SNMP Agent Server
;  PAGENT                   ; Policy Agent Server
;  PORTMAP                  ; Portmap Server (SUN 3.9)
;  PORTMAP JOBNAME PORTMAP1 ; USS Portmap Server (SUN 4.0)
;  RXSERVE                  ; Remote Execution Server
;
; --------------------------------------------------------------------
;
; PORT: Reserves a port for specified job names
;
;   - A port that is not reserved in this list can be used by any user.
;     If you have TCP/IP hosts in your network that reserve ports
;     in the range 1-1023 for privileged applications, you should
;     reserve them here to prevent users from using them.
;     The RESTRICTLOWPORTS option on TCPCONFIG and UDPCONFIG will also
;     prevent unauthorized applications from accessing unreserved
;     ports in the 1-1023 range.
;
;   - A PORT statement with the optional keyword SAF followed by a
```

```
;     1-8 character name can be used to reserve a PORT and control
;     access to the PORT with a security product such as RACF.
;     For port access control, the full resource name for the security
;     product authorization check is constructed as follows:
;     EZB.PORTACCESS.sysname.tcpname.safname
;     where:
;       EZB.PORTACCESS is a constant
;       sysname is the MVS system name (substitute your sysname)
;       tcpname is the TCPIP jobname (substitute your jobname)
;       safname is the 1-8 character name following the SAF keyword
;
;     When PORT access control is used, the TCP/IP application
;     requiring access to the reserved PORT must be running under a
;     USERID that is authorized to the resource. The resources
;     are defined in the SERVAUTH class.
;
;     For an example of how the SAF keyword can be used to enhance
;     security, see the definition below for the FTP data PORT 20
;     with the SAF keyword. This definition reserves TCP PORT 20 for
;     any jobname (the *) but requires that the FTP user be permitted
;     by the security product to the resource:
;     EZB.PORTACCESS.sysname.tcpname.FTPDATA in the SERVAUTH class.
;
;   - The BIND keyword is used to force a generic server (one that
;     binds to the IPv4 INADDR_ANY address, or the IPv6 unspecified
;     address, in6addr_any) to bind to the specific IP address that
;     is specified following the BIND keyword. This capability could
;     be used, for example, to allow z/OS UNIX telnet and telnet
;     3270 servers to both bind to TCP port 23.
;     The IP address that follows bind must be in IPv4 (dotted
;     decimal) or IPv6 (colon-hexadecimal) format and may be
;     any valid address for the host including VIPA and dynamic
;     VIPA addresses.
;
;   The special jobname of OMVS indicates that the PORT is reserved
;   for any application with the exception of those that use the Pascal
;   API.
;
;   The special jobname of * indicates that the PORT is reserved
;   for any application, including Pascal API socket applications.
;
;   Jobname may be 1 - 8 characters including wildcards. The following
;   wildcards are supported:
;   - An asterisk (*) can be used in any position to indicate zero or
;     more unspecified characters.
;   - A question mark (?) can be used in any position to indicate
;     a single unspecified character.
;
;   The special jobname of RESERVED indicates that the PORT is
;   blocked. It will not be available to any application.
;
;   GUIDELINE: When IPSECURITY is enabled, UDP ports 500 and 4500
;   should either be reserved for IKED (if it is in use) or should
;   be marked RESERVED.
;
;   TIP:  The PORT statement can also be used to control application
;   access to unreserved ports by configuring PORT entries where the
;   port number is replaced by the keyword UNRSV.
;
PORT
     7 UDP MISCSERV            ; Miscellaneous Server - echo
     7 TCP MISCSERV            ; Miscellaneous Server - echo
     9 UDP MISCSERV            ; Miscellaneous Server - discard
     9 TCP MISCSERV            ; Miscellaneous Server - discard
    19 UDP MISCSERV            ; Miscellaneous Server - chargen
    19 TCP MISCSERV            ; Miscellaneous Server - chargen
    20 TCP *  NOAUTOLOG        ; FTP Server
;   20 TCP *  NOAUTOLOG SAF FTPDATA ; FTP Server
    21 TCP FTPD1               ; FTP Server
    23 TCP TN3270              ; Telnet 3270 Server
;   23 TCP INETD1 BIND 9.67.113.3 ; z/OS UNIX Telnet server
   111 TCP PORTMAP             ; Portmap Server (SUN 3.9)
   111 UDP PORTMAP             ; Portmap Server (SUN 3.9)
;  111 TCP PORTMAP1            ; Unix Portmap Server (SUN 4.0)
;  111 UDP PORTMAP1            ; Unix Portmap Server (SUN 4.0)
   123 UDP SNTPD               ; Simple Network Time Protocol Server
   135 UDP LLBD                ; NCS Location Broker
   161 UDP OSNMPD              ; SNMP Agent
   389 TCP LDAPSRV             ; LDAP Server
   443 TCP HTTPS               ; http protocol over TLS/SSL
   443 UDP HTTPS               ; http protocol over TLS/SSL
;  500 UDP IKED                ; CS IKE daemon
```

```
   512 TCP RXSERVE             ; Remote Execution Server
   514 TCP RXSERVE             ; Remote Execution Server
;  512 TCP * SAF OREXECD       ; z/OS UNIX Remote Execution Server
;  514 TCP * SAF ORSHELLD      ; z/OS UNIX Remote Shell Server
;  515 TCP LPSERVE             ; LPD Server
;  515 TCP AOPLPD              ; Infoprint LPD Server
   520 UDP OMPROUTE            ; OMPROUTE Server (IPv4 RIP)
   521 UDP OMPROUTE            ; OMPROUTE Server (IPv6 RIP)
   750 TCP MVSKERB             ; Kerberos
   750 UDP MVSKERB             ; Kerberos
   751 TCP ADM@SRV             ; Kerberos Admin Server
   751 UDP ADM@SRV             ; Kerberos Admin Server
; 1700 TCP PAGENT NOAUTOLOG    ; Policy Agent pagentQosListener port
; 1701 TCP PAGENT NOAUTOLOG    ; Policy Agent pagentQosCollector port
  3000 TCP CICSTCP             ; CICS Socket
  3389 TCP MSYSLDAP            ; LDAP Server for Msys
; 4159 TCP NSSD                ; CS NSS daemon
; 4500 UDP IKED                ; CS IKE daemon
;16310 TCP PAGENT NOAUTOLOG    ; Policy Agent server listener port
;
; ----------------------------------------------------------------------
;
; PORTRANGE: Reserves a range of ports for specified jobnames.
;
;   In a common INET (CINET) environment, the port range indicated by
;   the INADDRANYPORT and INADDRANYCOUNT in your BPXPRMxx parmlib member
;   should be reserved for OMVS.
;
;   The special jobname of OMVS indicates that the PORTRANGE is reserved
;   for ANY z/OS UNIX socket application.
;
;   The special jobname of * indicates that the PORTRANGE is reserved
;   for any socket application, including Pascal API socket
;   applications.
;
;   Jobname may be 1 - 8 characters including wildcards. The following
;   wildcards are supported:
;   - An asterisk (*) can be used in any position to indicate zero or
;     more unspecified characters.
;   - A question mark (?) can be used in any position to indicate
;     a single unspecified character.
;
;   The special jobname of RESERVED indicates that the PORTRANGE is
;   blocked. It will not be available to any application.
;
;   The SAF keyword is used to restrict access to the PORTRANGE to
;   authorized users. See the use of SAF on the PORT statement above.
;
;
;   PORTRANGE 2000 1000 TCP RESERVED
;   PORTRANGE 3000  500 TCP APPL1*
;   PORTRANGE 4000 1000 TCP OMVS
;   PORTRANGE 4000 1000 UDP OMVS
;   PORTRANGE 5000 6000 TCP * SAF RANGE1
;
; ----------------------------------------------------------------------
;
; SACONFIG: Configures the SNMP TCP/IP subagent
;   The SACONFIG statement specified in this sample prevents the
;   activation of the TCP/IP subagent.  If you want the
;   TCP/IP subagent started during stack initialization, either
;   remove the sample statement, or respecify the statement as
;   follows, supplying a COMMUNITY and AGENT parameter value:
;
;     SACONFIG ENABLED COMMUNITY communityname AGENT portnum
;
;   If you remove the sample statement, the TCP/IP subagent will be
;   started with a COMMUNITY value of 'public', and an
;   AGENT port value of 161.
;
SACONFIG DISABLED
;
```

Figure 46. Example of reserved port number definitions

The following list describes the statements that are shown in Figure 46 on page 261. For more information about any of these statements, see z/OS Communications Server: IP Configuration Reference. For information specific to IPv6 support, see z/OS Communications Server: IPv6 Network and Application Design Guide.

AUTOLOG

Use AUTOLOG to list the procedure names that should start when the TCPIP address space starts. It is also used to supply a timeout value for detecting hung procedures at TCP/IP initialization time. The timeout value is the time TCP/IP should allow for a procedure to come down when, at startup, it is still active and TCP/IP is attempting to AUTOLOG the procedure again. A hung procedure is active to MVS, but is not listening on the socket that is reserved for it via the PORT statement. When AUTOLOG detects a hung task, TCP/IP checks every 10 seconds (until the timeout value has expired) to see if the procedure has come down. If the procedure comes down during one of these 10 second intervals, it is restarted. If the procedure is still active when the time interval specified by the timeout value expires, then TCP/IP cancels and restarts the procedure.

The AUTOLOG statement shown in Figure 46 on page 261 has a timeout value of five minutes.

In the first AUTOLOG statement the FTP Server shows FTPD JOBNAME FTPD1. This means when the TCPIP address space starts, the FTPD procedure will be started via the MVS START FTPD command. Because FTPD forks a child process that actually listens on PORT 21, the autolog task verifies that FTPD1 is listening on port 21.

Similarly, when the TCPIP address space starts, the autolog task starts the remaining 10 tasks.

Unless the tasks in the AUTOLOG list are in the PORT reservation list, the autolog task does not check for hung tasks every five minutes.

Note:

1. If you run multiple TCP/IP address spaces, ensure that the second address space AUTOLOG list does not cancel the procedures of the first. In those cases, an installation might require different procedure names for the servers for each address space. For more information about multiple stacks, see “Port management overview” on page 46.
2. You can use the AUTOLOG statement to automatically start generic servers in a single stack environment, but you should be careful using the AUTOLOG statement to start generic servers in a multiple stack environment. Instead, you could use an operations automation software package (IBM and other vendors provide these) to start generic servers automatically. For a list of generic servers provided by TCP/IP, see “Generic servers in a CINET environment” on page 48.

For those procedures that require parameters to be used on the MVS START command, there is a PARMSTRING option.

You should delay the start of AUTOLOG procedures that require AT-TLS services by specifying the optional DELAYSTART parameter with the TTLS subparameter on the AUTOLOG entry for the procedure. If you specify this parameter and subparameter, the procedure will start after the Policy Agent has installed the AT-TLS policy and AT-TLS services are available.

You should delay the start of AUTOLOG procedures that bind to a dynamic VIPA by specifying the optional DELAYSTART parameter with the DVIPA subparameter on the AUTOLOG entry for the procedure. If you specify this parameter and subparameter, the procedure will not start until the TCP/IP stack has joined the sysplex group and processed the dynamic VIPA configuration.

For a procedure that will bind to a dynamic VIPA and that requires AT-TLS services, you should specify DELAYSTART DVIPA TTLS. When more than one DELAYSTART subparameter is specified, all of the processing steps defined for those subparameters must complete before the procedure is started.

For more information, see z/OS Communications Server: IP Configuration Reference.

PORT

Use the PORT statement to do the following configuration:

- Reserve ports for different jobs and optionally limit access to these ports by user ID

  Use PORT to reserve ports for different jobs and to prevent rogue applications from taking ports intended for specific servers, such as port 21, which is needed by FTP. For each port entry, the port number, protocol, and procedure name are specified. The first port entry shows port 7 UDP reserved for the miscellaneous echo server for procedure MISCSERV. Similarly, port 7 of TCP is also reserved for the same server. In this example, six ports are reserved for the miscellaneous server.

NOAUTOLOG can be specified, as in the port 20 TCP * in Figure 46 on page 261. In this way, the port is reserved for an OMVS forked task so that the FTP server can fork tasks to port 20 as each FTP user logs in.

Use the DELAYACKS and NODELAYACKS options to allow an installation to delay their acknowledgments so they can be combined with data to be sent to foreign hosts. Unless a performance reason is needed, DELAYACKS should be used to delay the transmission of acknowledgments.

Use the SHAREPORT parameter or the SHAREPORTWLM parameter when reserving a port to be shared across multiple TCP listeners. This is not valid for UDP. To understand how these PORT statement parameters are used, see z/OS Communications Server: IP Configuration Reference. Use the Netstat ALL/-A report to determine whether port sharing is being used for a TCP listener. If port sharing is being used, this report indicates which type is being used.

Typically, reserving a port for a specific job name is sufficient. If the port must instead be reserved for a specific user ID or a set of user IDs, use the SAF keyword to specify the name of a SAF resource to be associated with the port. The user ID associated with the application that attempts to bind to the port must be permitted to the SAF resource.

The BIND keyword is used to force a generic server (one that binds to INADDR_ANY or in6addr_any) to bind to the specific IP address that is specified following the BIND keyword. This capability could be used, for example, to enable the z/OS UNIX Telnet and TN3270E Telnet servers to both bind to TCP port 23 on different IP addresses. The IP address that follows BIND can be any valid address for the host, including VIPA and dynamic VIPA addresses. The address supplied can be either an IPv4 address (in dotted decimal format) or an IPv6 address (in colon-hexadecimal format). IPv4-mapped IPv6 addresses are not supported. For multiple servers to bind to the same port with this function, the IP address for each server must be unique.

RESERVED indicates that the port is not available for use by any user.

- Control application access to unreserved ports

  You can also use the PORT statement to control application access to unreserved ports by configuring one or more PORT statements in which the port number is replaced by the keyword UNRSV. The UNRSV keyword refers to any unreserved port (any port number that has not been reserved by a PORT or PORTRANGE statement). If you configure the RESTRICTLOWPORTS parameter on the TCPCONFIG or UDPCONFIG profile statement, PORT UNRSV statements for the corresponding protocol control access only to unreserved ports above port 1023. If you do not configure the RESTRICTLOWPORTS parameter, PORT UNRSV statements control access to all unreserved ports in the range 1-65535. The type of access that is controlled by the PORT UNRSV statement is specified (explicitly or by default) by the WHENBIND or WHENLISTEN keywords.

  If you configure one or more PORT UNRSV statements for a protocol, access is unconditionally denied to any application that explicitly binds to an unreserved port and that does not match the protocol and job name on any of the configured PORT UNRSV statements. Applications that explicitly bind to an unreserved port and that do match the protocol and job name on a PORT UNRSV statement are allowed to access the unreserved port, unless the access is restricted by the SAF or DENY keywords. If the SAF keyword is specified, the user ID associated with the application that attempts to access the port must be permitted to the specified SAF resource. If the DENY keyword is specified, access is unconditionally denied.

  For UDP sockets, the access permission is checked when an unreserved port is specified on an explicit bind. The WHENBIND keyword is the only access option that is allowed for UDP ports.

  For TCP sockets, access can be controlled when an unreserved port is specified on an explicit bind (WHENBIND) or when a listen is issued on a user-specified port that was not reserved for the application (WHENLISTEN).

  For more information about controlling access to unreserved ports, see “Port access control” on page 137. For more information about the PORT statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

PORTRANGE
PORTRANGE is a statement used to reserve a range of ports for specified job names.

SACONFIG
SACONFIG is the statement used to configure the information about the SNMP TCP/IP subagent. The AGENT keyword on this statement is used to specify the port number to be used when the TCP/IP subagent connects to the SNMP agent. Omission of this statement causes TCPIP to assume the default value of SACONFIG ENABLED COMMUNITY public AGENT 161. For more information on the SACONFIG profile statement, see z/OS Communications Server: IP Configuration Reference.

Setting up the System Authorization Facility server access authorization class (optional)

The TCP/IP address space uses the server access authorization (SERVAUTH) class of the System Authorization Facility (SAF) to protect TCP/IP resources from unauthorized access. The use of SERVAUTH may be optional and is available in degrees so that installations can pick and choose the access needed. Installations may be able to choose to use one, all, or none of the protections provided by SERVAUTH. The customization described here is completely optional when using the IBM security product RACF. Non-IBM security products might require customization. A template of the commands and all other SAF commands appears in SEZAINST(EZARACF). See Chapter 3, “Security,” on page 123 for more detailed information.

Configuring the local host table (optional)

You can set up the local host table to support local host name resolution. If you use the local host table for this purpose, your socket applications will be able to resolve only names and IP addresses that appear in your local host table.

If you need to resolve host names outside your local area, you can configure the resolver to use a domain name server (see the NSINTERADDR statement). If you use a domain name server, you do not need to set up any host definitions in your resolver configuration, but you may still do so.

If you have configured your resolver to use a name server, it will always try to do so, unless your TCP/IP C/C++ API applications were written with a RESOLVE_VIA_LOOKUP symbol in the source code. You can also configure the resolver to use only a local host table by specifying LOOKUP LOCAL in the TCPIP.DATA configuration file. For both cases, all name resolution calls will always use a local host table. This is probably not a technique you will see for standard socket applications, but it may be a technique you could find useful for when you develop your own socket applications or for testing changes before they are placed in your name server.

It might be a good idea to have a local host table available for the resolver to use if the name server is not reachable. If the name server does not respond to name resolution requests, the resolver tries to use the local host table. If the name server is reachable but returns a negative reply for a name resolution request, the resolver tries to resolve the name using the local host table, if such a file is present.

Assume you try to resolve the host name `friendly` and your DOMAINORIGIN is `my.house.com`, the resolver sends a query to the name server for `friendly.my.house.com`. If the name server returns a negative reply (the name is not registered), the resolver looks into the local host table for an entry of `friendly.my.house.com` and, if not found, for an entry of `friendly`.

Due to the flexibility of the Domain Name System, it is recommended you use a domain name server. If you set up a small IP network, the simplicity of the local host table approach might be preferable.

The following types of local host table can be used:

- HOSTS.LOCAL

  HOSTS.LOCAL is used only for IPv4 requests.
- /etc/hosts

  /etc/hosts is used only for IPv4 requests.
- ETC.IPNODES and /etc/ipnodes

ETC.IPNODES and /etc/ipnodes can be used for IPv6 requests, and for IPv4 requests when COMMONSEARCH is coded in the resolver setup statements.

Creating HOSTS.LOCAL site host table

The site host table is generated from the *hlq*.HOSTS.LOCAL data set. This data set contains descriptions of local host entries in the HOSTS format. HOSTS.LOCAL can contain only IPv4 addresses. A sample HOSTS.LOCAL data set is created during installation. This information describes how to update the sample *hlq*.HOSTS.LOCAL data set and use it to generate the two data sets, *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO, which function as your site table.

HOST entries

One line of the *hlq*.HOSTS.LOCAL data set is used for each distinct host and ends with four colons (::::). The maximum length of the line is 512 characters. Each host can have multiple IP addresses and multiple names. The line for each host has the following three essential fields, which are separated by colons:

- The keyword HOST.
- A list of IP addresses for that host, which are separated by commas. You can specify a maximum of six IP addresses per HOST entry.
- A list of fully qualified names for that host, which are separated by commas. The first name is the official host name, or CNAME (canonical name), and any additional host names after the CNAME are alias names. You can specify a maximum of 20 host names. Only the first six host names are used in the *hlq*.HOSTS.ADDRINFO data set. All 20 host names are used in the *hlq*.HOSTS.SITEINFO data set.

For example, if you have two local hosts, LOCAL1 (IP addresses 192.6.77.4 and 192.8.4.1) and LOCAL2 (with an alias LOCALB and IP address 192.6.77.2), append the following lines to the *hlq*.HOSTS.LOCAL data set:

```
HOST : 192.6.77.4, 192.8.4.1 : LOCAL1 ::::
HOST : 192.6.77.2 : LOCAL2, LOCALB ::::
```

Notes:

1. The maximum length for a host that is allowed in the HOST tables is 24 characters.
2. If the HOST entry has more than the maximum of 6 IP addresses for one host name or more than 20 host names for a single IP address, the MAKESITE command completes successfully with no error message issued.

NET and GATEWAY entries

The NET and GATEWAY statements are not used by TCP/IP for z/OS applications. However, some socket calls require the NET entries. If your programs do not need the NET and GATEWAY statements, delete them before invoking MAKESITE.

Sample HOSTS.LOCAL data set (HOSTS)

The following sample HOSTS.LOCAL data set is provided in SEZAINST(HOSTS):

```
; HOSTS.LOCAL
; -----------
; COPYRIGHT = NONE.
;
; The format of this file is documented in RFC 952, "DoD Internet
; Host Table Specification".
;
; The format for entries is:
;
; NET : ADDR : NETNAME :
; GATEWAY : ADDR, ALT-ADDR : HOSTNM : CPUTYPE : OPSYS : PROTOCOLS :
; HOST : ADDR, ALT-ADDR : HOSTNM, NICKNM : CPUTYPE : OPSYS : PROTOCOLS :
```

```
;
; Where:
;   ADDR, ALT-ADDR = IP address in decimal, e.g., 26.0.0.73
;   HOSTNM, NICKNM = the fully qualified host name and any nicknames
;   CPUTYPE = machine type (PDP-11/70, VAX-11/780, IBM-3090, C/30, etc.)
;   OPSYS = operating system (UNIX, TOPS20, TENEX, VM/SP, etc.)
;   PROTOCOLS = transport/service (TCP/TELNET,TCP/FTP, etc.)
;   : (colon) = field delimiter
;   :: (2 colons) = null field
; *** CPUTYPE, OPSYS, and PROTOCOLS are optional fields.
;
;   MAKESITE does not allow continuation lines, as described in
;   note 2 of the section "GRAMMATICAL HOST TABLE SPECIFICATION"
;   in RFC 952.  Entries should be specified on a single line of
;   up to a maximum of 512 characters per line.
;
;
; Note:  The NET and GATEWAY statements are not used by the TCP/IP for
;        MVS applications.  However, some socket calls require the NET
;        entries.  For better performance, if your programs do not need
;        the NET and GATEWAY statements, delete them before running
;        the MAKESITE program.
;
;
HOST : 9.67.43.100 : NAMESERVER ::::
HOST : 9.67.43.126 : RALEIGH ::::
HOST : 129.34.128.245, 129.34.128.246 : YORKTOWN, WATSON ::::
;
NET  : 9.67.43.0 : RALEIGH.IBM.COM :
;
GATEWAY : 129.34.0.0 : YORKTOWN-GATEWAY ::::
;
```

Using MAKESITE

Because many servers and commands allocate *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO, it is important not to overwrite or delete these data sets while TCP/IP is running. To avoid disrupting any active users, use an *HLQ*=parm that is different than your active *hlq*. This allows you to swap names (by renaming the old HOSTS data sets and then renaming the new HOSTS data sets) without starting and stopping TCP/IP.

Use MAKESITE as a TSO command or in a batch job to generate new *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO data sets. The parameters are the same for either a TSO command or a batch job invocation of MAKESITE. See z/OS Communications Server: IP System Administrator's Commands for more information.

After you make changes to your *hlq*.HOSTS.LOCAL data set, you must generate and install new *hlq*.HOSTS.SITEINFO and *hlq*.HOSTS.ADDRINFO data sets.

Guidelines: Use ETC.IPNODES (in the style of etc/ipnodes) as the preferred alternative to the generated local hosts tables from MAKESITE for the following reasons:

- No imposed 24 character restriction on host names.
- No imposed restriction on the first eight characters of the host names having to be unique. However, there are certain applications that require the first eight characters to be unique, such as Network Job Entry (NJE).
- Closely resembles that of other TCP/IP platforms, and eliminates the MAKESITE requirement of file post-processing.
- Allows configuration of both IPv4 and IPv6 addresses.
- Only one file is managed for the system, and that all the APIs can use the same single file. The COMMONSEARCH statement in the resolver setup file can be used to reduce IPv6 and IPv4 searching to a single search order, as well as to reduce the z/OS UNIX and native MVS environments to a single search order.

For more information on the use of IPNODES by the resolver to locate IPv4 and IPv6 addresses and sitenames, see z/OS Communications Server: IPv6 Network and Application Design Guide.

For the search orders used in locating the local host tables, see “Configuration files for TCP/IP applications” on page 25. For a description of the use of IPNODES by the resolver to locate IPv4 and IPv6 addresses and sitenames when the resolver setup statement COMMONSEARCH is specified, see “IPv6/common search order” on page 780 and “IPv6/common search order” on page 784.

Creating /etc/hosts

The `/etc/hosts` file can be defined as follows:

- The maximum line length is 256 characters. If a line is greater than 256 characters, it is truncated to 256 characters and processed. If Trace Resolver is active, a warning message is issued.
- The line starts with an IP address, followed by a blank, followed by host names. Host names are separated by one or more blanks. The first host name that follows an IP address is the official host name, or CNAME (canonical name). Any additional host names that follow the CNAME are alias host names.
- Only IPv4 addresses are supported.
- Each IP address can have up to 35 host names. A host name can have up to 35 IP addresses.
- The values for the host name must conform to the following rules:
  - Maximum of 128 characters.
  - Must contain one or more tokens that are separated by a period.
  - Each token must be at least one character and less than 64 characters.
  - The first character in each token must be a letter (A-Z or a-z) or number (0-9).
- A comment is indicated by the # or ; character.

For the search orders used in locating `/etc/hosts`, see “Configuration files for TCP/IP applications” on page 25.

Creating ETC.IPNODES and /etc/ipnodes

The ETC.IPNODES and `/etc/ipnodes` file can be defined as follows:

- z/OS UNIX files can be in any directory. The maximum line length that is supported is 256 characters. If a line is greater than 256 characters, it is truncated to 256 characters and processed. If Trace Resolver is active, a warning message is issued. If you need more than 256 characters for the IP address and the associated host names, you can code more lines with the same IP address as follows:

```
Address1 Hostname1
Address1 Hostname2
Address1 Hostname3
```

- MVS data sets must be partitioned organization (PO) or sequential (PS), RECFM=F or RECFM=FB, a logical record length (LRECL) 56 - 256, and have any valid blocksize (BLKSIZE) for fixed block.
- It can contain IPv4 and IPv6 addresses, but not IPv4 mapped addresses. Each IP address can have up to 35 host names. A host name can have up to 35 IP addresses.
- The line starts with an IP address, followed by a blank, followed by host names. Host names are separated by one or more blanks. The first host name that follows an IP address is the official host name, or CNAME (canonical name). Any additional host names that follow the CNAME are alias host names.
- The values for the host name must conform to the following rules:
  - Maximum of 128 characters.
  - Must contain one or more tokens that are separated by a period.
  - Each token must be at least one character and less than 64 characters.
  - The first character in each token must be a letter (A-Z or a-z) or number (0-9).
- A comment is indicated by the # or ; character.

The sample IPNODES file that is provided by z/OS Communications Server follows. It can be found as member EZBREIPN (alias IPNODES) in SEZAINST.

```
;
;  IBM z/OS Communications Server
;  SMP/E distribution name: EZBREIPN
;
;  5694-A01 (C) Copyright IBM Corp. 2002.
;  Licensed Materials - Property of IBM
;
;  Function: Sample ETC.IPNODES file
;
;  The file contains the Internet Protocol (IP) host names
;  and addresses for the local host and other hosts in the
;  Internet network.
;  This file is used to resolve a name into an address (that is, to
;  translate a host name into its Internet address) or resolve
;  an address into a name.
;
;  Comments begin with a # or ; character and continue until the
;  end of the line.
;
;  The following statement defines the Internet Protocol (IP) name
;  and address of the local host and specifies the names and
;  addresses of remote hosts. The maximum line length support is
;  256 characters
;
;  Entries in the hosts file have the following format:
;
;  Address HostName
;
;  Address HostName1 HostName2 HostName3 ..... HostName35
;
;
;  Address: is an IP address, it can be IPV4 or IPV6 address.
;           Note: IPv4-mapped IPv6 address is not allowed.
;
;  HostName: the length of the hostname is up to 128 characters,
;            and each IP address can have up to 35 hostnames.
;
;
   9.67.43.100 NAMESERVER
   9.67.43.126 RALEIGH
   9.67.43.222 HOSTNAME1.RALEIGH.IBM.COM
   129.34.128.245 YORKTOWN WATSON
   1::2 TESTIPV6ADDRESS1
   1:2:3:4:5:6:7:8 TESTIPV6ADDRESS2
;
```

For the search orders used in locating ETC.IPNODES and `/etc/ipnodes`, see “Configuration files for TCP/IP applications” on page 25.

Verifying your configuration

At this point, your configuration files have been updated.

To verify a configuration, start the TCP/IP address space and ensure that you see the following message:

```
EZB6473I TCP/IP STACK FUNCTIONS INITIALIZATION COMPLETE
```

If the message is not displayed, the messages issued by the TCP/IP address space should describe why TCP/IP did not start.

The following message will be generated after EZB6473I above and indicates complete initialization of TCP/IP and extended services:

```
EZD1314I TCP/IP AND EXTENDED SERVICES ARE NOW INITIALIZED FOR STACK: stackname
```

If the message is not displayed, see Diagnosing TCP/IP stack initialization problems in z/OS Communications Server: IP Diagnosis Guide.

Verifying TCPIP.DATA statement values in the native MVS environment

To display which TCPIP.DATA statement values are being used and where they are being obtained from, use Trace Resolver output. You can obtain Trace Resolver output at your TSO screen by issuing the following TSO commands:

```
alloc f(systcpt) dsn(*)
READY
netstat up
READY
free f(systcpt)
READY
```

Tips:

- When you direct Trace Resolver output to a TSO terminal, define the screen size to be only 80 columns wide. Otherwise, trace output is difficult to read.
- You can also use the Resolver CTRACE function to collect Trace Resolver output.

For more information about using the Trace Resolver facility and interpreting its output, or about using the Resolver CTRACE function to collect Trace Resolver output, see z/OS Communications Server: IP Diagnosis Guide.

Verifying TCPIP.DATA statement values in the z/OS UNIX environment

To display which TCPIP.DATA statement values are being used and where they are being obtained from, use Trace Resolver output. You can obtain Trace Resolver output by issuing the following z/OS UNIX shell commands:

```
export RESOLVER_TRACE=stdout
netstat -u
set -A RESOLVER_TRACE
```

Tip: You can also use the Resolver CTRACE function to collect Trace Resolver output.

For more information about using the Trace Resolver facility and interpreting its output, or about using the Resolver CTRACE function to collect Trace Resolver output, see z/OS Communications Server: IP Diagnosis Guide.

Verifying PROFILE.TCPIP

You can use the following commands to verify many configuration values that are specified in the TCP/IP profile:

- Netstat commands

  You can use the Netstat DEvlinks/-d command to verify the physical network and hardware definitions. You can use the Netstat CONFIG/-f command to verify operating characteristics.

- DISPLAY TCPIP,,OSAINFO command

  You can use this command to verify that the OSA-Express QDIO configuration was successfully accepted by the OSA-Express QDIO feature. The command output is a view from the OSA-Express QDIO feature of the data subchannel device for the interface.

For information about the syntax and output of these commands, see z/OS Communications Server: IP System Administrator's Commands.

Verifying interfaces with Ping and Traceroute

The Ping and Traceroute commands can be used in the TSO and z/OS UNIX environments to verify adapters or interfaces attached to the z/OS host. For information about the syntax and output of the commands, see z/OS Communications Server: IP System Administrator's Commands.

Given that your PROFILE.TCPIP file contains the interfaces of your installation and that the TCPIP.DATA file contains the correct TCPIPJOBNAME, the TCPIP address space is configured and you can go on to configuring routes, servers, and so on.

Verifying local name resolution with TESTSITE

Use the TESTSITE command to verify that the *hlq*.HOSTS.ADDRINFO and *hlq*.HOSTS.SITEINFO data sets can correctly resolve the name of a host, gateway, or net. For more information on the TESTSITE command, see z/OS Communications Server: IP System Administrator's Commands.

Verifying PROFILE.TCPIP and TCPIP.DATA using HOMETEST

Use the HOMETEST command to verify the HOSTNAME, DOMAINORIGIN, SEARCH, and NSINTERADDR TCPIP.DATA statements. HOMETEST will use the resolver to obtain the IP addresses assigned to the HOSTNAME and compare them to the HOME list specified in PROFILE.TCPIP. A warning message will be issued if any HOSTNAME IP addresses are missing from the HOME list.

Activate Trace Resolver to obtain detailed information about how the HOSTNAME is resolved to IP addresses. The information will also include what TCPIP.DATA data set names were used. This can be done by issuing the following TSO command before running HOMETEST. The detailed information will be displayed on your TSO screen.

```
allocate dd(systcpt) da(*)
```

Issue the following TSO command after HOMETEST to turn off Trace Resolver output.

```
free dd(systcpt)
```

Tip: You can also use the Resolver CTRACE function to collect Trace Resolver output.

If you do not have Trace Resolver turned on before you run HOMETEST, the following information is displayed:

```
hometest

Running IBM MVS TCP/IP CS V1R6 TCP/IP Configuration Tester

FTP.DATA file not found. Using hardcoded default values.

TCP Host Name is: MVS026

Using Name Server to Resolve MVS026
The following IP addresses correspond to TCP Host Name: MVS026
9.67.113.58

The following IP addresses are the HOME IP addresses defined in PROFILE.TCPIP:
9.67.113.58
127.0.0.1

All IP addresses for MVS026 are in the HOME list!

Hometest was successful - all Tests Passed!
```

For more information about using the Trace Resolver facility and interpreting its output, or about using the Resolver CTRACE function to collect Trace Resolver output, see z/OS Communications Server: IP Diagnosis Guide.

Verifying your X Window System installation (Optional)

Note: You cannot verify your X Window System until after routing and DNS setup.

Support is provided for two versions of the X Window System and the corresponding Motif. The current support, provided as part of the base IP support in z/OS Communications Server, is for X Window System Version 11 Release 6 and Motif Version 1.2. Support for X Window System Version 11 Release 4 and Motif Version 1.1 is available as feature HIP614X.

Verifying the X Window X11R4 System installation

X Window X11R4 System is installed with the other target libraries. The macro or headers go into the target library data set SEZACMAC. To verify the installation of the X Window System:

1. Specify your workstation IP address by adding a record (such as the following record) to your XWINDOWS.DISPLAY data set.

   ```
   royal.csc.ibm.com:0.0
   ```

   In this example,

   ```
   royal.csc.ibm.com:0.0
   ```

   is the name of the host running the X Window System server.

   Note: No leading blanks are allowed in this record.
2. On the workstation running the X Window System server, issue an XHOST command specifying the name of your MVS system.
3. Run the program with the XSAMP1 command.

Verifying the X Window X11R6 System installation

To verify the installation of the X Window X11R6 System:

1. Ensure that a host (the workstation) with an X Window System server that supports X11R6 is properly configured and reachable by the MVS system. From the workstation, use Telnet to access the MVS system, and open a z/OS UNIX shell on the MVS system.
2. From the z/OS UNIX shell, export the DISPLAY environment variable using either the network name or the qualified IP address of the workstation as shown in the following example:

   ```
   export DISPLAY=royal.csc.ibm.com:0.0
   ```

   In this example, *royal.csc.ibm.com* is the name of the workstation running the X Window System server. The display is indicated by *:0.0*, and is specified this way in almost all cases.
3. Authorize the MVS system to access the workstation by executing the XHOST command, and specify either the name of the MVS system or a plus sign (+) as shown in the following example.

   ```
   xhost +
   ```

   Note: The *+* option turns off security for this workstation and allows any X client to display here.
4. The sample X clients are shipped in the directory */usr/lpp/tcpip/X11R6/Xamples/demos*. Change into this directory. There are four sample program directories, xsamp1, xsamp2, xsamp3, and pexsamp. Change to the xsamp1 directory. Verify that there are files named Makefile and xsamp1.c, and then execute the following command:

   ```
   make
   ```
5. Execute the program using the following command:

   ```
   xsamp1
   ```
6. The z/OS UNIX shell should block as another window is opened. Verify the workstation is displaying a new window. The xsamp1 client displays a blank window for 60 seconds and then exits, taking its window with it. The z/OS UNIX shell should no longer be blocked.

Customizing TCP/IP messages

The messages for every application provided by TCP/IP are compiled and linked with the application and are stored in an internal message repository. Some of the applications also store their messages in

message catalogs or message data sets. You can modify these message catalogs or message data sets to translate the messages to another language or to customize them for your installation.

Customizing message catalogs

Applications in Table 16 on page 272 use message catalogs that are created by using the **gencat** utility. The message catalogs are installed in the /usr/lpp/tcpip/lib/nls/msg/C/ directory in the z/OS UNIX file system.

Table 16. TCP/IP message catalogs

Application	**Message catalogs**
Autolog TCP/IP configuration	cfmsg.cat
certbundle command (z/OS UNIX)	certbundlemsg.cat
Communications Server SMTP (CSSMTP) application	ezamlcat.cat ezamlrpy.cat
Defense Manager daemon (DMD)	dmdmsg.cat
Digital Certificate Access Server (DCAS)	dcasm.cat
DISPLAY TCPIP,,NETSTAT operator command NETSTAT command (TSO) **netstat**/**onetstat** command (z/OS UNIX)	netmsg.cat netmsg6.cat
DISPLAY TCPIP,,SYSPLEX command VARY TCPIP,,SYSPLEX command	spxmsg.cat
dnsdomainname command (z/OS UNIX) **domainname** command (z/OS UNIX) **host** command (z/OS UNIX) **hostname** command (z/OS UNIX)	ezahomsg.cat ezaitmsg.cat
FTP client FTP server	ftpdmsg.cat ftpdrply.cat
IKE daemon(IKED)	ike_log.cat
ipsec command (z/OS UNIX)	ipsecmsg.cat
Load Balancing Advisor (LBA) Automated domain name registration (ADNR)	lcadvm.cat
MODDVIPA utility	xfdvpm.cat
Motif V1.1 XWindows X11R4	Mrm12.cat Uil12.cat xm12.cat X11R6.cat

Table 16. TCP/IP message catalogs (continued)	
Application	**Message catalogs**
Motif V1.2 XWindows X11R6	Mrm21.cat xm21.cat Uil21.cat X11R66.cat
Network security services (NSS) server	nssdmsg.cat
nssctl command (z/OS UNIX)	nssctlmsg.cat
Omproute	omprdmsg.cat
orexec/**rexec** command (z/OS UNIX) **orsh**/**rsh** command (z/OS UNIX) REXEC command (TSO) RSH command (TSO)	rexcmsg.cat
pasearch command (z/OS UNIX) Policy Agent(Pagent)	pagentm.cat
PING command (TSO) **ping** command (z/OS UNIX)	pingmsg.cat
popper (z/OS UNIX)	popper_msg.cat
PORTMAP (z/OS UNIX) RPC Rpcbind	msrpcbin.cat msrpcgen.cat msrpclib.cat
REXECD server (z/OS UNIX)	rexdmsg.cat
RSHD server (z/OS UNIX)	rshdmsg.cat
Simple Network Time Protocol daemon (SNTPD)	sntpdmsg.cat
SNMP Agent (OSNMPD)	snmpdmsg.cat
snmp command (z/OS UNIX)	snmpclim.cat
SNMP Network SLAPM2 subagent (nslapm2)	pagtsmsg.cat
SNMP TN3270E Telnet subagent	ezbtsmsg.cat
Syslog daemon (syslogd)	syslogd.cat
TCP/IP SNMP subagent	subamsg.cat
Telnet server (z/OS UNIX)	tnmsgs.cat
TIMED daemon	timedmsg.cat
TRACERTE command (TSO) **traceroute** command (z/OS UNIX)	trtemsg.cat
Traffic regulation manager daemon (TRMD)	trmdm.cat
Trap forwarder daemon (TRAPFWD)	trapfwd.cat

Table 16. TCP/IP message catalogs (continued)	
Application	**Message catalogs**
trmdstat command (z/OS UNIX)	trmdstat.cat

Message format

Use the z/OS UNIX **dspcat** command to display the message text from a message catalog. The messages are in the following format:

```
index_num "message text"
```

The *index_num* value is the identifier that the application uses to access the message text.

The *message text* value might include conversion characters for the variable fields that are converted when the message is printed or displayed and control characters that affect the message format. The conversion characters start with a percent sign (%) and the control characters start with a backslash (\). These are all standard notations for the C language print function. The messages might also contain comments, which start with /* and end with */.

In the following simulated message, the control character \n forces a new line to print. The string variables %1$s, %2$s, and %3$s are converted in the order that they are passed from the application.

```
1 "EZZ2350I MVS TCP/IP NETSTAT %1$s       TCPIP Name: %2$s        %3$s\n"
```

Rules for modifying messages

The general rule for customizing or translating messages is to change only the text portion of the message.

- Do not change the *index_num* identifier. This part of the message has a specific meaning. If you change it, the application might not function correctly.
- Do not change the conversion characters for the variables. These characters indicate that the application is passing data, the type of data that the application is passing, and the appropriate way to display or print this data. For example, do not change or delete %1$s and %2$d in the following message:

  ```
  475 "EZA1644I Recfm: %1$s  lrecl: %2$d\n"
  ```

  If you change or delete the conversion characters, results are unpredictable.
- You can reorder the sequence of variables within a message when the variables are defined by conversion characters with the format %*n*$, where *n* is an integer. For example, you can reverse the conversion characters when translating the following message:

  ```
  Before:
  480 "EZY9999I Command %1$s received from user %2$s\n"

  After:
  480 "EZY9999I Utilisador %2$s envio instrucion %1$s\n"
  ```

  The result would be `EZY9999I Utilisador MANNY envio instrucion FTP` instead of `EZY9999I Command FTP received from user MANNY`.

  You cannot reorder the sequence of variables in the following message:

  ```
  EZY9999I Command $s received from user $s
  ```

Steps for creating a modified message catalog

Some applications store their messages in message catalogs. You can modify these message catalogs to translate the messages to another language or to customize them for your installation.

Before you begin

Review the following information:

- "Message format" on page 274
- "Rules for modifying messages" on page 274

You also must know:

- The commands that are indicated in these steps are z/OS UNIX commands; do all of these steps from the z/OS UNIX shell.
- These steps use a message catalog that is called sample.cat, and assume that the application code and catalog are at the correct levels. If you customize a catalog, IBM Service personnel might require that you use the shipped level of the catalog to re-create and diagnose a reported problem.

Procedure

Perform the following steps to create a modified message catalog:

1. Copy the current z/OS UNIX catalog to a backup file to ensure that you have a copy of the original catalog.

   ```
   cp  /usr/lpp/tcpip/lib/nls/msg/C/sample.cat  /usr/lpp/tcpip/lib/nls/msg/C/sample.cat.backup
   ```

2. Using the z/OS UNIX **dspcat** command, convert the editable backup catalog that you just copied.

   ```
   dspcat -t -g  /usr/lpp/tcpip/lib/nls/msg/C/sample.cat.backup  > /tmp/sample.cat.copy
   ```

   The file sample.cat.copy is the file that you will update to preserve the timestamp and customize the messages.

3. Change the first line in the catalog from a comment to a z/OS UNIX **gencat** command $timestamp directive.

   This enables the original timestamp to be imbedded in the new catalog when the catalog is built. In some cases, when an application opens a message catalog, it checks the timestamp at the top of the catalog to ensure that it matches the application timestamp. If the timestamps are different, the application does not use the message catalog. You must ensure that the new message catalog contains the timestamp from the original message catalog by performing the following steps:

   a. Edit the file to be updated.

      ```
      oedit /tmp/sample.cat.copy
      ```

   b. Change the first line comment to add a **gencat** command $timestamp directive to preserve the timestamp when the directory is built. The first line in the file is similar to the following line:

      ```
      The time stamp of catalog /usr/lpp/tcpip/lib/nls/msg/C/sample.cat.backup is: 2010 095
      20:30 UTC
      ```

      Replace the leading text on the line with the $timestamp directive as follows:

      ```
      $timestamp 2010 095 20:30 UTC
      ```

      If you omit this step and the original line is left unchanged in the catalog, when you attempt to generate a catalog from this file, you will see a message similar to the following message:

      ```
      FSUM5108 gencat: Invalid message number.
      ```

   c. Save the file.

4. Update the catalog (/tmp/sample.cat.copy) with any local modifications and save the file.
5. Build a new and customized catalog by using the z/OS UNIX **gencat** command.

   ```
   gencat   /tmp/sample.cat   /tmp/sample.cat.copy
   ```

   The correct response is the following message:

   ```
   FSUM5105 gencat: Message catalog generated normally.
   ```

   Tip: Verify that your changes are correct before invoking the **gencat** command. Some errors cause the **gencat** command to fail, but other errors might not be apparent until the message is displayed.
6. Browse the new catalog and verify that the timestamp from step "3" on page 275 matches what is in the file.

   ```
   obrowse /tmp/sample.cat
   ```

   The first record contains the time stamp (*yyyy ddd hh:mm* UTC). For example:

   ```
   ...2010 095 20:30 UTC
   ```
7. Replace the z/OS UNIX catalog, which is shipped with the release, with the updated catalog that you created in step "5" on page 276.

   ```
   cp  /tmp/sample.cat /usr/lpp/tcpip/lib/nls/msg/C/sample.cat
   ```

What to do next

For more information about the **dspcat** and **gencat** utilities, see z/OS UNIX System Services Command Reference.

Customizing message data sets

The following table shows the servers that have external messages, the DD statement used, and the name of the message data set delivered with the system:

Server	DD statement	Data set
SNMP Query Engine	//MSSNMPMS	SEZAINST(MSSNMP)
MISC Server	//MSMISCSR	SEZAINST(MSMISCSR)

The procedures for these servers have a special DD statement that point to the external message data set. If you are going to override the internal messages and use external customized messages, you need to remove the comment from the appropriate DD statement and ensure it points to the correct data set.

Message text

The message text might include conversion characters for the variable fields that are converted when the message is printed or displayed, and control characters that affect the message format. The conversion characters start with a percent sign (%) and the control characters start with a backslash (\). These are all standard notations for the C language print function. The messages might also contain comments which start with /* and end with */.

In the following simulated message, the control character \n forces a new line to print and the string variables, represented by %s, are converted in the order they are passed from the program.

```
9999 I "Command %s received from user %s\n"
```

Message format

Figure 47 on page 277 explains the syntax for TCP/IP message IDs on the host:

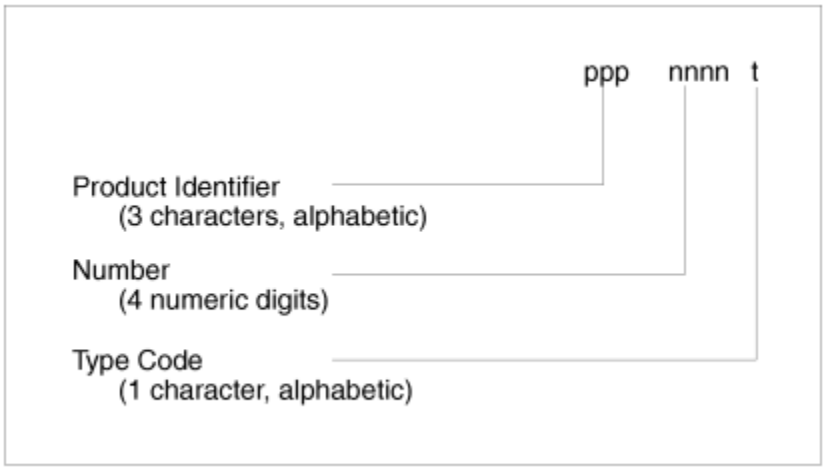


Figure 47. Syntax for TCP/IP message IDs

The *product identifiers* (ppp) for TCP/IP are EZA, EZB, EZD, EZY, EZZ, and SNM. The *number* (nnnn) indicates a unique 4-digit numeric value assigned to the message by product. The *type code* (t) indicates the severity assigned to the message.

Rules for customizing the messages

The general rule for customizing or translating messages is to change only the text portion of the message.

- Do not change the MARGIN, PRODUCT, and COMPONENT definitions at the top of the data set. These are required definitions for the program. For example, these entries at the top of the MISC server message data set should not be changed:

  ```
  MARGINS(1,72)
  PRODUCT EZA
  COMPONENT MSC
  ```

- Do not change the message numbers and the severity code. These parts of the message have specific meaning; if you change them the program may not work correctly.
- Do not change the conversion characters. These indicate that the program is passing data, the type of data it is passing, and the appropriate way to display or print this data. For example, do not change or delete %s and %d in the following message:

  ```
  4059 I "Connecting to agent %s on DPI port %d\n"
  ```

- You can reorder the variables that are passed in the message. For example, you can reverse the order of the two string variables that are passed when translating a message by specifying the new order of the arguments in parentheses following the message text:

  Before: `9999 I "Command %s received from user %s\n"`

  After: `9999 I "Utilisador %s envio instrucion %s\n"(2, 1)`

  The result would be `EZY9999I Utilisador MANNY envio instrucion FTP` instead of `EZY9999I Command FTP received from user MANNY`.

- Watch for any program parameters or keywords that might be in the message text. In most cases, you should not translate them.

  For example, in the following message, `'active'` is a keyword used in the gateway definition and should not be translated:

  ```
  3974 E   "First two elements must be 'active' for active gateway\n"
  ```

Chapter 6. Routing

This information explains the steps that are required to configure your TCP/IP stack for dynamic, static, or policy-based routing, and explains how to verify the configuration. If you understand your entire network configuration, after you read this information you can do the following tasks:

- Configure dynamic, static, or policy-based routing
- Use the Ping command to ping a remote host by IP address
- Use the Traceroute command to determine the path that will be taken to reach a particular destination using static or dynamic routing
- Use the Netstat command to display a TCP/IP stack's routing information
- Use DISPLAY commands to display dynamic routing information

Guideline: The definition or modification of an installation's routing configuration should not be performed without a complete understanding of the entire network design.

Routing terminology

"General terms" on page 279 describes some of the more common IP routing-related terms and concepts. If you need more detailed information, see *Routing in the Internet* by Christian Huitema.

General terms

Autonomous system (AS)
A group of routers that are exchanging routing information through a common routing protocol. A single AS can represent many IP networks.

Dynamic routes
IP layer routing table entries that are dynamically managed and can automatically change in response to network topology changes. For IPv4, these routes are managed by a routing daemon. For IPv6, these routes can be managed by a routing daemon, and they can also be learned by listening to router advertisement messages received from routers as part of the router discovery protocol.

Exterior Gateway Protocol (EGP)
A routing protocol that is spoken by routers that belong to different autonomous systems when those routers are configured to share routing information between autonomous systems. This topic does not discuss exterior gateway routing.

Interior Gateway Protocol (IGP)
A routing protocol that is spoken by routers that belong to the same autonomous system. Each AS has a single IGP. A separate AS within a network can be running a different IGP.

Main route table
An IPv4 or IPv6 route table that is populated by using static routes and dynamic routes. These route tables have the name EZBMAIN. A TCP/IP stack has one main IPv4 route table and one main IPv6 route table. When policy-based routing is not in use, the main route tables contain all of the routes that a TCP/IP stack uses when it is making routing decisions. When policy-based routing is in use, policies can be configured to use the main route tables in routing decisions when no route is found in a policy-based route table.

Policy-based route table
A route table that is configured for use by policy-based routing. A TCP/IP stack can have zero or more policy-based route tables. A policy-based route table can be defined by using a flat file that is parsed by the Policy Agent, or by using the IBM Configuration Assistant for z/OS Communications Server. A policy-based route table definition can contain static routes, dynamic routing parameters for controlling the scope of dynamic routes that are added to the table, or both. Policy rules and actions can then be defined to indicate, for given types of traffic, which route tables the TCP/IP stack is to use when it is making routing decisions.

© Copyright IBM Corp. 2000, 2022

Policy-based routing (PBR)
A technique that is used to make routing decisions that are based on policies that are defined by the network administrator. Policy-based routing selects a route for outbound traffic from a set of policy-based route tables, and optionally the main route table, according to the policy defined for the traffic.

Replaceable static routes
IPv4 static routes that can be replaced by OMPROUTE, or IPv6 static routes that can be replaced by OMPROUTE or by routes that are learned by listening to router advertisement messages received from routers as part of the router discovery protocol.

Router
A device or host that interprets protocols at the IP layer and forwards datagrams on a path towards their correct destination.

Routing
The process that is used in an IP network to deliver a datagram to the correct destination.

Routing daemon
A server process that manages the IP route table.

Static routes
IP layer routing table entries that are manually configured (by using the BEGINROUTES configuration statement) and do not change automatically in response to network topology changes, except when:

- The change is due to an ICMP redirect (if not disabled).
- A dynamic routing protocol learns a route to a destination configured as a replaceable static route.

Interior Gateway Protocols

An interior gateway protocol (IGP) is a dynamic route update protocol used between routers that run on TCP/IP hosts within a single autonomous system. The routers use this protocol to exchange information about IP routes.

Some of the more common interior gateway protocols are:

Routing Information Protocol (RIP)
RIP uses a distance vector algorithm to calculate the best path to a destination based on the number of hops in the path. RIP has several limitations. Some of the limitations which exist in RIP Version 1 are resolved by RIP Version 2.

RIP Version 2
RIP Version 2 extends RIP Version 1. Among the improvements are support for multicasting and variable subnetting. Variable subnetting allows the division of networks into variable size subnets. For example, one route can represent addresses from 9.1.1.0 through 9.1.1.255 (the 9.1.1.0/255.255.255.0 subnet) while another can represent addresses from 9.2.0.0 through 9.2.255.255 (the 9.2.0.0/255.255.0.0 subnet).

IPv6 RIP
IPv6 RIP uses the same distance vector algorithm used by RIP to calculate the best path to a destination. It is intended to allow routers to exchange information for computing routes through an IPv6-based network.

Open Shortest Path First (OSPF)
OSPF uses a link state or shortest path first algorithm. OSPF's most significant advantage compared to RIP is the reduced time needed to converge after a network change. In general, OSPF is more complicated to configure than RIP and might not be suitable for small networks.

IPv6 OSPF
IPv6 OSPF also uses a link state or shortest path first algorithm to calculate the best path to a destination. IPv6 OSPF has the same advantages and more complicated configuration compared to IPv6 RIP, like OSPF compared to RIP.

Table 17. Interior Gateway Protocol characteristics

Feature	RIP-1	RIP-2	IPv6 RIP	OSPF	IPv6 OSPF
Algorithm	Distance vector	Distance vector	Distance vector	Shortest path first	Shortest path first
Network load (1)	High	High	High	Low	Low
CPU processing requirement (1)	Low	Low	Low	High	High
IP network design restrictions	Many	Some	Some	Virtually none	Virtually none
Convergence time	Up to 180 seconds	Up to 180 seconds	Up to 180 seconds	Low	Low
Multicast supported (2)	No	Yes	Yes	Yes	Yes
Multiple equal-cost routes	No (3)	No(3)	No(3)	Yes	Yes

Notes:

1. Depends on network size and stability.
2. Multicast saves CPU cycles on hosts that are not interested in certain periodic updates, such as OSPF link state advertisements or RIP-2 routing table updates. Multicast frames are filtered out either in the device driver or directly on the interface card if this host has not joined the specific multicast group.
3. RIP in OMPROUTE allows multiple equal-cost routes only for directly connected destinations over redundant interfaces. See “Use of static routing with OMPROUTE” on page 357.

Route selection algorithm

Whether static, dynamic, or policy-based routing is used, the algorithm used by the IP layer to select a route from a route table is the same. Route selection occurs in the following order:

1. If a route exists to the destination address (a host route), it is chosen.
2. If no host route exists to the destination address, the route chosen depends upon the version of IP being used:
   - For IPv4:
     a. If subnet, network, or supernetwork routes exist to the destination, the route with the most specific network mask (the mask with the most bits on) is chosen.
     b. If the destination is a multicast destination and a multicast default route exists, that route is chosen.
   - For IPv6, if prefix routes exist to the destination, the route with the most specific prefix is chosen.
3. Default routes are chosen when no other route exists to a destination.

Multiple equal-cost routes are allowed for static, dynamic, and policy-based routing. Table 17 on page 281 and “Multiple equal-cost routes” on page 295 provide additional information about the use of multiple equal-cost routes.

In the absence of policy-based routing, the IP layer routes traffic by searching the main route table for the most specific route known, using the selection order described. If policy-based routing is being used, the

IP layer routes traffic according to the policy defined for the traffic. For more information about how the IP layer routes traffic when policy-based routing is being used, see "Policy-based routing" on page 349.

The sample network

Figure 48 on page 282 and Figure 49 on page 283 show network diagrams that depict a sample network. This sample network is used in the following topics as the configuration of static, dynamic, and policy-based routing is described:

- "IPv4 static routing" on page 283
- "IPv6 static routing" on page 285
- "IPv4 dynamic routing using OMPROUTE" on page 288
- "IPv6 dynamic routing using OMPROUTE" on page 291
- "Policy-based routing" on page 349

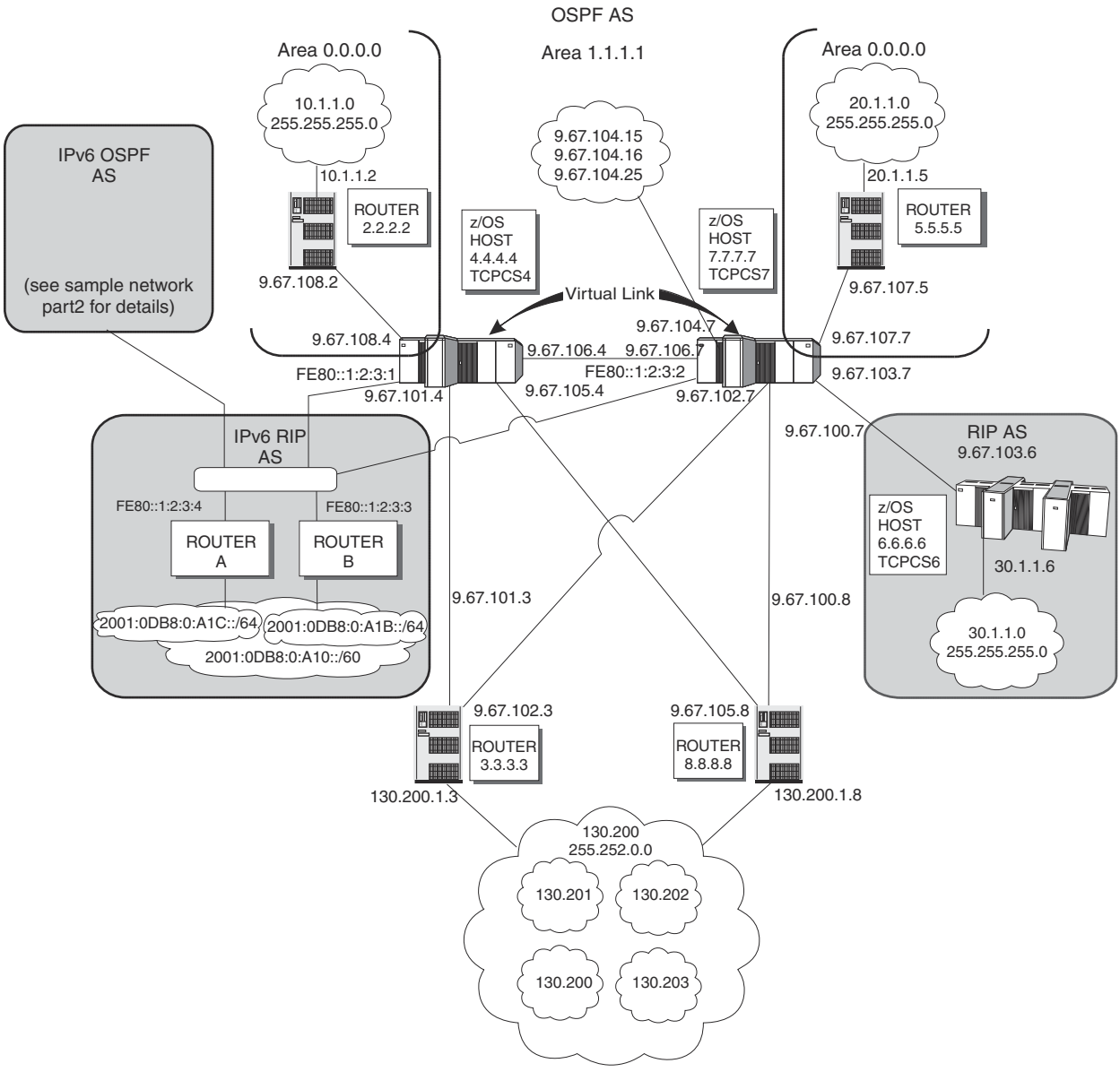


Figure 48. Sample network part 1

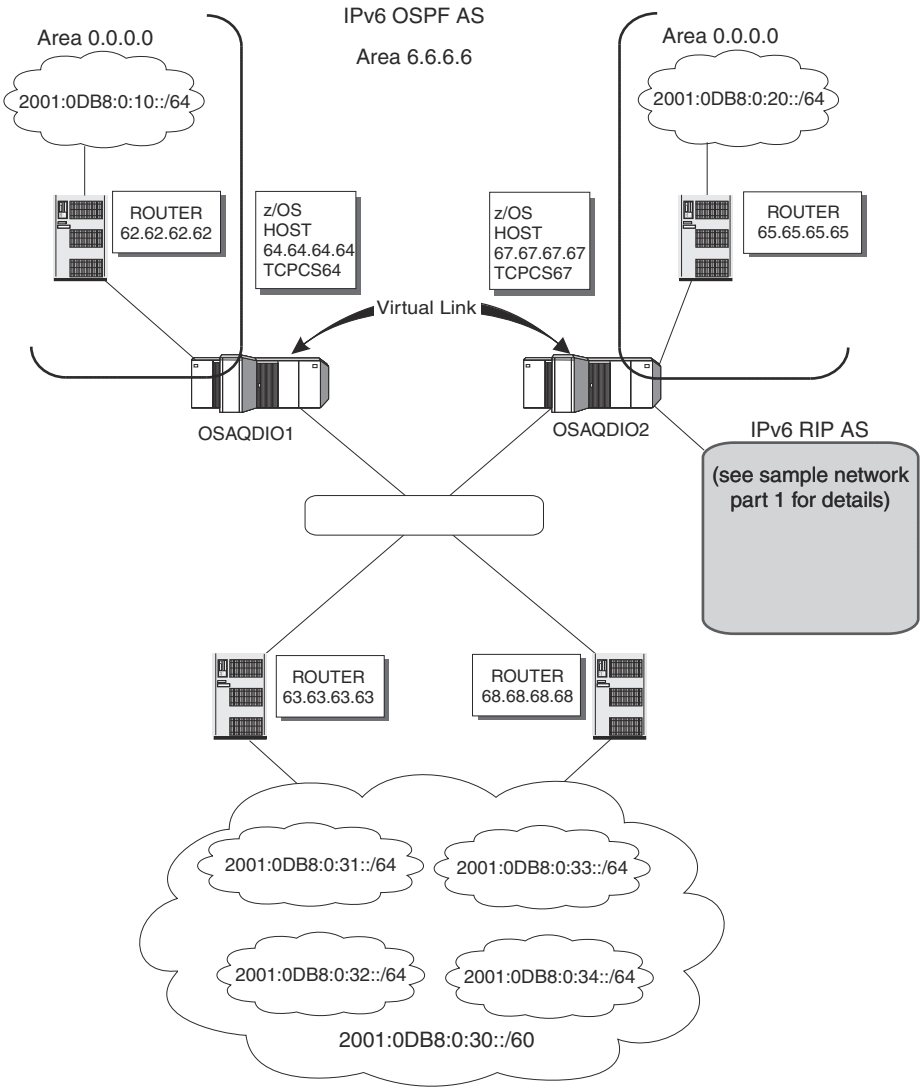


Figure 49. Sample network part 2, IPv6 OSPF AS

In this sample network, TCPCS4 and TCPCS7 are both performing as OSPF area border routers between OSPF areas 0.0.0.0 and 1.1.1.1. TCPCS64 and TCPCS67 are both performing as OSPF area border routers between IPv6 OSPF areas 0.0.0.0 and 6.6.6.6. TCPCS7 is performing as an AS boundary router between the OSPF AS and the RIP AS. TCPCS67 is performing as an AS boundary router between the IPv6 OSPF AS and the IPv6 RIP AS. TCPCS4 and TCPCS7 have interfaces to both the IPv4 network and the IPv6 network.

IPv4 static routing

Static routing requires that routes are configured manually for each router or destination; this is a significant reason system administrators avoid this technique if given a choice. Static routing has the disadvantage that network reachability is not dependent on the state of the network itself. If a destination is down or unreachable over that statically configured route, the static routes remain in the routing table, and traffic continues to be sent toward that destination without success.

To minimize network administrator tasks, configuration of static routes is to be avoided, especially in a large network. However, certain circumstances make static routing more appropriate. For example, static routes can be used:

- To define a default route or a route that is not being advertised within a network using a dynamic routing protocol
- To replace exterior gateway protocols when trying to avoid:

  - The cost of routing protocol traffic between ASs
  - Complex routing policies
- In conjunction with a routing daemon to provide backup routes when the daemon cannot find a dynamic route to the destination. In this case, the static route must be configured as replaceable.

If static routing is used, only the PROFILE.TCPIP data set has to be updated with the BEGINROUTES statement.

The only ways to modify static routes are:

- Replace the routing table using the VARY TCPIP,,OBEYFILE command
- Use incoming ICMP Redirect packets
- Use ICMP Must Fragment packets
- If a static route is defined on a BEGINROUTES statement as being replaceable, it can be replaced if a dynamic route is discovered by OMPROUTE. This is the only way that a static route can be replaced by a dynamic route.

For more information on the VARY TCPIP,,OBEYFILE command, see z/OS Communications Server: IP System Administrator's Commands. For more information on the IPCONFIG statement, and the IGNOREREDIRECTS and PATHMTUDISC parameters for the IPCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

The first BEGINROUTES statement in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command, replaces all static routes in the TCP/IP stack routing table, including those destination addresses specified in the BSDROUTINGPARMS section of the PROFILE.TCPIP. Subsequent statements within the same data set append to the routing table.

Every interface must have an IP address to transmit or receive packets. Along with the IP address, each interface must have a subnet mask associated with it for routing purposes. The combination of the address and mask will yield the subnet that the interface belongs to and also determines the broadcast address for the interfaces. For interfaces not involved in dynamic routing, the subnet mask can be specified in one of the following ways:

- On the BSDROUTINGPARMS statement in PROFILE.TCPIP.
- On the INTERFACE statement in the OMPROUTE configuration file, if running OMPROUTE. Also, if running OMPROUTE, OMPROUTE will always override the stack's BSDROUTINGPARMS and set the subnet mask for every interface it is aware of, even if you did not code the interface to OMPROUTE. Therefore, it is highly recommended that you either specify every interface to OMPROUTE with the correct subnet mask, or configure OMPROUTE to ignore undefined interfaces.

If you do not specify the subnet mask by one of these methods, the subnet mask is determined by the stack using the stack routing table. Specifying the subnet mask rather than letting it default is highly recommended.

For point-to-point links, such as CTC, a destination address is used to document the other end of a point-to-point connection. For interfaces not involved in dynamic routing, this can be specified in one of the following ways:

- On the BSDROUTINGPARMS statement in PROFILE.TCPIP.
- On the INTERFACE statement in the OMPROUTE configuration file, if running OMPROUTE. Also, if running OMPROUTE, OMPROUTE will always override the stack's BSDROUTINGPARMS and set the destination address for every point-to-point link it is aware of, even if you did not code the interface to OMPROUTE. Therefore, it is highly recommended that you either specify every interface to OMPROUTE with the correct destination address, or configure OMPROUTE to ignore undefined interfaces.

If you do not specify the destination address by one of these methods, the stack determines the address by using the stack routing table. Specifying the destination address rather then letting it default is highly recommended.

Tip: You can use the IBM Health Checker for z/OS to check whether the total number of indirect routes in a TCP/IP stack routing table exceeds a maximum threshold. When this threshold is exceeded, OMPROUTE

and the TCP/IP stack can experience high CPU consumption from routing changes. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide and IBM Health Checker for z/OS: User's Guide.

Replaceable static routes

Because replaceable static routes are intended to be last-resort routes, TCP/IP attempts to use them only if no dynamic routes to the destination are available.

If a non-replaceable static route fails validation, even if the reason for the failure is transient like gateway unreachable, the definition for the non-replaceable static route is discarded. However, if a replaceable static route fails validation for a transient reason, the definition of the route is retained. When there are no dynamic routes to the destination, TCP/IP periodically tries to add the replaceable static route to the routing table. Because of these periodic attempts, multiple EZZ4333I messages might be seen. Retries are performed no more often than every 30 seconds, and only as long as there are no active routes to the destination in the routing table, and only if at least one new route has been added to the routing table since the last retry. Retries are terminated as soon as a valid route to the destination is installed into the routing table, whether it is dynamic, static, or replaceable static.

For a discussion of things to consider when using IPv4 static routing and OMPROUTE together on the same TCP/IP stack, see “Use of static routing with OMPROUTE” on page 357.

IPv6 static routing

Static routing requires that routes are configured manually for each router or destination; this is a significant reason system administrators avoid this technique if given a choice. Static routing has the disadvantage that network reachability is not dependent on the state of the network itself. If a destination is down, or unreachable through a statically configured route, the static routes remain in the routing table and traffic continues to be sent toward that destination without success.

To minimize network administrator tasks, configuration of static routes is to be avoided, especially in a large network. However, certain circumstances make static routing more appropriate. For example, static routes can be used:

- To define a route that will not be learned dynamically from OMPROUTE or from router advertisements received from routers as part of the router discovery protocol
- In conjunction with dynamic routes to provide backup routes when a dynamic route to the destination cannot be found. In this case, the static route must be configured as replaceable.

If static routing is used, only the PROFILE.TCPIP data set has to be updated with BEGINROUTES statements. The only ways to modify static routes are:

- Replace the routing table using the VARY TCPIP,,OBEYFILE command
- Use incoming ICMPv6 redirect packets
- If a static route is defined on a BEGINROUTES statement as being replaceable, it can be replaced by a dynamic route

Note that the first BEGINROUTES statement in PROFILE.TCPIP, or in the data set referenced by a VARY TCPIP,,OBEYFILE command, replaces all static routes in the TCP/IP stack routing table. Subsequent statements within the same data set append to the routing table.

Rule: If you use static routes and want to honor ICMPv6 redirect messages (that is, you do not code IPCONFIG6 IGNOREREDIRECTS), then you must code the first hop address using the link-local address of the router. This is required because all redirect messages are sent using the router's link-local address, and if the source address of the redirect message does not match the address of the first hop in the routing table, the redirect message will be ignored.

For more information on the VARY TCPIP,,OBEYFILE command, see z/OS Communications Server: IP System Administrator's Commands. For more information on the IPCONFIG6 statement, and the IGNOREREDIRECTS parameter on the IPCONFIG6 statement, see z/OS Communications Server: IP Configuration Reference.

Tip: You can use the IBM Health Checker for z/OS to check whether the total number of indirect routes in a TCP/IP stack routing table exceeds a maximum threshold. When this threshold is exceeded, OMPROUTE and the TCP/IP stack can experience high CPU consumption from routing changes. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide and IBM Health Checker for z/OS: User's Guide.

Replaceable static routes

Because replaceable static routes are intended to be last-resort routes, TCP/IP attempts to use them only if no dynamic routes to a destination are available. If a non-replaceable static route fails validation, even if the reason for the failure is transient (for example, gateway unreachable), the definition for the non-replaceable static route is discarded. However, if a replaceable static route fails validation for a transient reason, the definition of the route is retained. When there are no dynamic routes to the destination, TCP/IP periodically tries to add the replaceable static route to the routing table. Because of these periodic attempts, multiple EZZ4348I messages might be seen. Attempts occur every 30 seconds at the most, if there are no active routes to the destination in the routing table and at least one new route was added to the routing table since the last attempt. Attempts end when a valid route to the destination is installed into the routing table, whether it is a dynamic, static, or replaceable static route.

For a discussion of things to consider when using IPv6 static routing and IPv6 router advertisements together on the same TCP/IP stack, see “Use of IPv6 static routing with router advertisements” on page 357. For a discussion of things to consider when using IPv6 static routing and OMPROUTE together on the same TCP/IP stack, see “Use of static routing with OMPROUTE” on page 357.

Static routing configuration examples

The following topics illustrate static routing configuration examples.

z/OS TCPCS4

Static route statements for z/OS TCPCS4:

```
BEGINRoutes ;first BEGINRoutes in the profile
;
;Network/mask                 FirstHop      LinkName         PacketSize
Route 9.67.106.0/24           =             CTC4TO7          MTU 1500       ;route1
Route 9.67.105.0/24           =             CTC4TO8          MTU 1500       ;route2
Route 9.67.101.0/24           =             CTC4TO3          MTU 1500       ;route3
Route 9.67.108.0/24           =             CTC4TO2          MTU 1500       ;route4
Route 9.67.107.0/24           9.67.106.7    CTC4TO7          MTU 1500       ;route5
Route 7.7.7.7/32              9.67.106.7    CTC4TO7          MTU 1500       ;route6
Route 9.67.103.0/24           9.67.101.3    CTC4TO3          MTU 1500       ;route7
Route 9.67.103.0/24           9.67.106.7    CTC4TO7          MTU 1500       ;route8
Route 30.1.1.0/24             9.67.106.7    CTC4TO7          MTU 1500       ;route9
Route 10.1.1.0/24             9.67.108.2    CTC4TO2          MTU 1500       ;route10
Route 130.200.0.0/14          9.67.101.3    CTC4TO3          MTU 1500       ;route11
Route 130.200.0.0/14          9.67.105.8    CTC4TO8          MTU 1500       ;route12
Route 130.203.0.0/16          9.67.105.8    CTC4TO8          MTU 1500       ;route13
Route DEFAULT                 9.67.106.7    CTC4TO7          MTU 1500       ;route14
;
;Destination/PrefixLen        FirstHop      Interface        PacketSize
Route FE80::1:2:3:3/128       =             OSAQDIO46        MTU 5000 REPL  ;route15
Route FE80::1:2:3:4/128       =             OSAQDIO46        MTU 5000 REPL  ;route16
Route 2001:0DB8:0:A1B::/64    FE80::1:2:3:3 OSAQDIO46        MTU 5000 REPL  ;route17
Route 2001:0DB8:0:A1C::/64    FE80::1:2:3:4 OSAQDIO46        MTU 5000 REPL  ;route18
Route DEFAULT6                FE80::1:2:3:4 OSAQDIO46        MTU 5000 REPL  ;route19
EndRoutes
;
```

Notes:

1. In the BEGINROUTES block, the netmask can be specified by a /xx. This number, denoted by xx, represents the number of significant bits in the netmask. For example:

   ```
   /16 = 16 significant bits = 11111111 11111111 00000000 00000000 = 255.255.0.0
   ```

   For IPv6, you must specify the prefix length of the route using the /xxx notation.

2. For direct routes, use an equals symbol (=) for the first hop.

BSDROUTINGPARMS statements for z/OS TCPCS4:

```
BSDRoutingParms TRUE ; Shown only for completeness
; Linkname    MTU    Metric    Subnet Mask     Dest Address
  CTC4TO8     1500        0    255.255.255.0   0
  CTC4TO7     1500        0    255.255.255.0   0
  CTC4TO3     1500        0    255.255.255.0   0
  CTC4TO2     1500        0    255.255.255.0   0
  VIPA1A      1500        0    255.255.255.252 0
  EndBSDRoutingParms
;
```

z/OS TCPCS7

Static route statements for z/OS TCPCS7:

```
BEGINRoutes
;
;Network/mask              FirstHop      LinkName   PacketSize
Route 9.67.106.0/24        =             CTC7TO4   MTU 1500       ;route1
Route 9.67.100.0/24        =             CTC7TO8   MTU 1500       ;route2
Route 9.67.102.0/24        =             CTC7TO3   MTU 1500       ;route3
Route 9.67.103.0/24        =             CTC7TO6   MTU 1500       ;route4
Route 9.67.107.0/24        =             CTC7TO5   MTU 1500       ;route5
Route 4.4.4.4/32           9.67.106.4    CTC7TO4   MTU 1500       ;route6
Route 10.1.1.0/24          9.67.106.4    CTC7TO4   MTU 1500       ;route7
Route 20.1.1.0/24          9.67.107.5    CTC7TO5   MTU 1500       ;route8
Route 30.1.1.0/24          9.67.103.6    CTC7TO6   MTU 1500       ;route9
Route 130.200.0.0/14       9.67.100.8    CTC7TO8   MTU 1500       ;route10
Route 130.200.0.0/14       9.67.102.8    CTC7TO3   MTU 1500       ;route11
Route 130.203.0.0/16       9.67.102.3    CTC7TO3   MTU 1500       ;route12
Route DEFAULT              9.67.107.5    CTC7TO5   MTU 1500       ;route13
;
;Destination/PrefixLen     FirstHop      Interface  PacketSize
Route FE80::1:2:3:3/128    =             OSAQDIO76 MTU 5000 REPL  ;route14
Route FE80::1:2:3:4/128    =             OSAQDIO76 MTU 5000 REPL  ;route15
Route 2001:0DB8:0:A1B::/64 FE80::1:2:3:3 OSAQDIO76 MTU 5000 REPL  ;route16
Route 2001:0DB8:0:A1C::/64 FE80::1:2:3:4 OSAQDIO76 MTU 5000 REPL  ;route17
Route DEFAULT6             FE80::1:2:3:4 OSAQDIO76 MTU 5000 REPL  ;route18
EndRoutes
```

BSDROUTINGPARMS statements for z/OS TCPCS7:

```
BSDRoutingParms TRUE
; Linkname    MTU    Metric    Subnet Mask     Dest Address
  CTC7TO8     1500        0    255.255.255.0   0
  CTC7TO3     1500        0    255.255.255.0   0
  CTC7TO6     1500        0    255.255.255.0   0
  CTC7TO4     1500        0    255.255.255.0   0
  CTC7TO5     1500        0    255.255.255.0   0
  VIPA1A      1500        0    255.255.255.252 0
  EndBSDRoutingParms
;
```

The sample configuration has an IPv4 supernet route for 130.200.0.0. An IPv4 supernet route means that the netmask for the route is smaller than the class netmask. In this case, 130.200.0.0 is a class B address. The default netmask for class B is 255.255.0.0. The netmask used for this sample is 255.252.0.0, which is less than 255.255.0.0, hence making this a supernet route. In routing, the stack prefers a route that has the most bits in common. Therefore, the stack chooses a route in the following order:

1. If a route exists to the destination address (a host route), it is chosen.
2. At this point, the route chosen depends upon the version of IP being used:
   - For IPv4:
     a. If subnet, network, or supernetwork routes exist to the destination, the route with the most specific network mask (the mask with the most bits on) is chosen.

b. If the destination is a multicast destination and a multicast default route exists, that route is chosen.

- For IPv6, if prefix routes exist to the destination, the route with the most specific prefix is chosen.

3. Default routes are chosen when no other route exists to a destination.

For example, for TCPCS4 (and when trying to reach 130.200.0.0), route12 in the list is used, which is the supernet route 130.200.0.0 with mask 255.252.0.0. If applying the mask of that route, 255.252.0.0, to the destination IP address, 130.200.0.0, the result is 130.200.0.0, which is the IP address of this route. Now, when trying to reach destination 130.203.5.2, the stack would use route13 in the list, which is a network route for 130.203.00 with mask 255.255.0.0. If applying the mask of that route, 255.255.0.0, to the destination IP address, 130.203.5.2, the result is 130.203.00, which is the IP address of this route.

For TCPCS4, route7 and route8 are examples of equal cost multipath routes to get to 9.67.103.0 subnet. This means that TCPCS4 has two different routes to get to this destination. If IPCONFIG MULTIPATH is not enabled, then only route7 will be used as long as it is active. This is because the stack chooses the first route and ignores route8. If route7 becomes inactive, then the stack will switch and use route8. If MULTIPATH is enabled, then the stack will use both routes according to the MULTIPATH specification.

In the preceding example, all of the IPv4 links have a subnet mask of 255.255.255.0 because this is what is specified for the links in the BSDROUTINGPARMS. Therefore, to determine the broadcast addresses for link CTC4TO3, AND the IP Address, 9.67.101.4, and the subnet mask, 255.255.255.0, to yield the subnet for this link, 9.67.101.0. Then, OR the subnet, 9.67.101.0, with the complement of the subnet mask, 0.0.0.255. This determines that the broadcast address for this link is 9.67.101.255.

For TCPCS4, route15 and route16 would be selected to reach host FE80::1:2:3:3 and host FE80::1:2:3:4 respectively. Route17 and route18 would be selected to reach any IPv6 address that had the first 64 bits of 2001:0DB8:0:A1B and 2001:0DB8:0:A1C respectively. Route19 would be selected for any other IPv6 destination.

Rules:

1. All IPv4 IP addresses must follow Classless Inter-Domain Routing (CIDR) convention that requires the actual mask to be one or more on-bits followed by zero or more off-bits. On-bits cannot be followed by off-bits followed by on-bits. Therefore, a mask of 255.255.254.0 is valid (an actual mask of FFFFE00), but a mask of 255.255.253.0 is not valid (an actual mask of FFFFD00) because 253 is 11111101.
2. VIPA links or VIPA interfaces are not allowed on BEGINROUTES statements.
3. DEFAULT and DEFAULT6 routes are always indirect routes and therefore must always have a first hop address specified.

IPv4 dynamic routing using OMPROUTE

Daemon is a UNIX term for a background server process, and daemons are used for dynamic routing. For z/OS Communications Server IP, there is a multiprotocol routing daemon available. For IPv4, OMPROUTE supports the RIP Version 1, RIP Version 2, and OSPF routing protocols. You can send RIP Version 1 or RIP Version 2, but not both at the same time on a single interface. However, you can configure a RIP interface to receive both versions.

Guideline: If OROUTED was used in a prior release and the RIP protocol is still the preferred dynamic routing method in your host configuration, use OMPROUTE to provide RIP support.

For IPv4, OMPROUTE implements the OSPF protocol described in RFC 1583 (*OSPF Version 2*), the OSPF subagent protocol described in RFC 1850 (*OSPF Version 2 Management Information Base*), and the RIP protocols described in RFC 1058 (*Routing Information Protocol*) and in RFC 1723 (*RIP Version 2 - Carrying Additional Information*). It provides an alternative to the static TCP/IP gateway definitions. The MVS host running with OMPROUTE becomes an active OSPF or RIP router in an IP network. Either or both of these routing protocols can be used to dynamically maintain the host routing table. For example, OMPROUTE can detect when a route is created, is temporarily unavailable, or if a more efficient route exists. If both OSPF and RIP protocols are used simultaneously, OSPF routes are preferred over RIP routes to the same destination.

Tip: You can use the IBM Health Checker for z/OS to check whether the total number of indirect routes in a TCP/IP stack routing table exceeds a maximum threshold. When this threshold is exceeded, OMPROUTE and the TCP/IP stack can experience high CPU consumption from routing changes. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide and IBM Health Checker for z/OS: User's Guide.

Open Shortest Path First

Open Shortest Path First (OSPF) is classified as an Interior Gateway Protocol (IGP). This means that it distributes routing information between routers belonging to a single autonomous system (AS), a group of routers all using a common routing protocol. The OSPF protocol is based on link-state or shortest path first (SPF) technology. It has been designed expressly for the TCP/IP Internet environment, including explicit support for IP subnetting and the tagging of externally derived routing information.

OSPF performs the following tasks:

Multiple routes
Provides support for up to 16 equal-cost routes.

Authentication
Provides for the authentication of routing updates.

IP multicast
Uses IP multicast when sending or receiving the updates.

Allows network grouping
Allows sets of networks to be grouped together. Such a grouping is called an area. The topology of an area is hidden from the rest of the autonomous system. This method of hiding information enables a significant reduction in routing traffic. Also, routing within the area is determined only by the area's own topology, lending the area protection from bad routing data. An area is a generalization of an IP subnetted network.

IP subnet configuration
Enables the flexible configuration of IP subnets. Each route distributed by OSPF has a destination and mask. Two different subnets of the same IP network number may have different sizes (that is, different masks). This is commonly referred to as variable length subnetting. A packet is routed to the best (longest or most specific) match. Host routes are considered to be subnets whose masks are all ones (0xFFFFFFFF).

Authenticate OSPF protocol exchanges
Can be configured such that all OSPF protocol exchanges are authenticated. This means that only trusted routers can participate in the autonomous system's routing. A single authentication scheme is configured for each physical link. This enables some links to use authentication while others do not.

OSPF is a dynamic routing protocol. It quickly detects topological changes in the AS (such as router interface failures) and calculates new loop-free routes after a period of convergence. This period of convergence is short and involves a minimum of routing traffic as compared to the RIP protocol.

In a link-state routing protocol, each router maintains a database describing the autonomous system's topology. Each individual piece of this database is a particular router's local state (for example, the router's usable interfaces and reachable neighbors). The router distributes its local state throughout the autonomous system by flooding.

To generate routes, all routers run the exact same algorithm, in parallel. From the topological database, each router constructs a tree of shortest paths with itself as root. This shortest-path tree gives the route to each destination in the autonomous system. Externally derived routing information (for example, routes learned from the RIP protocol) appears on the tree as leaves. When multiple equal-cost routes to a destination exist, the routes (up to 16) are added to the TCP/IP stack's route table. The TCP/IP stack uses these equal-cost routes according to the multipath setting configured for the route table. “Multiple equal-cost routes” on page 295 provides additional information about the multipath setting configured for a route table and the use of multiple equal-cost routes.

Externally derived routing data (for example, routes learned from the RIP protocol) is passed transparently throughout the autonomous system. This externally derived data is kept separate from

the OSPF protocol's link state data. Each external route can also be tagged by the advertising router, but not by OMPROUTE, enabling the passing of additional information between routers on the boundaries of the autonomous system. OMPROUTE does pass tags created by others. For information on configuring OSPF, see “Steps for configuring OSPF and RIP (IPv4 and IPv6)” on page 305.

Routing Information Protocol

Routing Information Protocol (RIP) is an Interior Gateway Protocol (IGP) designed to manage a relatively small network. RIP is based on the Bellman-Ford or the distance-vector algorithm. RIP has many limitations and is not suitable for every TCP/IP environment. Before using the RIP function in OMPROUTE, read RFCs 1058 and 1723 to decide if RIP can be used to manage the routing tables of your network. For more information about RFCs 1058 and 1723, see Appendix G, “Related protocol specifications,” on page 1409.

RIP uses the number of hops, or hop count, to determine the best possible route to a host or network. The term hop count is also referred to as the metric. In RIP, a hop count of 16 means infinity, or that the destination cannot be reached. This limits the longest path in the network that can be managed by RIP to 15 gateways.

A RIP router broadcasts routing information to its directly connected networks every 30 seconds. It receives updates from neighboring RIP routers every 30 seconds and uses the information contained in these updates to maintain the routing table. If an update has not been received from a neighboring RIP router in 180 seconds, a RIP router assumes that the neighboring RIP router is down, sets all routes through that router to a metric of 16 (infinity), and stops using those routes when routing IP packets. If an update has still not been received from the neighboring RIP router after another 120 seconds, the RIP router deletes from the routing table all of the routes through that neighboring RIP router.

RIP Version 2 is an extension of RIP Version 1 and provides the following features:

Route Tags
The route tags are used to separate *internal* RIP routes (routes for networks within the RIP routing domain) from *external* RIP routes, which may have been imported from an EGP (external gateway protocol) or another IGP. OMPROUTE does not generate route tags, but preserves them in received routes and readvertises them when necessary.

Variable subnetting support
Variable length subnet masks are included in routing information so that dynamically added routes to destinations outside subnetworks or networks can be reached.

Immediate next hop for shorter paths
Next hop IP addresses, whenever applicable, are included in the routing information to eliminate packets being routed through extra hops in the network.

Multicasting to reduce load on hosts
IP multicast address 224.0.0.9, reserved for RIP Version 2 packets, is used to reduce unnecessary load on hosts which are not listening for RIP Version 2 messages. This support is dependent on interfaces that are multicast-capable.

Authentication for routing update security
Authentication keys can be configured for inclusion in outgoing RIP Version 2 packets. Incoming RIP Version 2 packets are checked against the configured keys.

Configuration switches for RIP Version 1 and RIP Version 2 packets
Configuration parameters allow for controlling which version of RIP packets are to be sent or received over each interface.

Supernetting support
The supernetting feature is part of Classless InterDomain Routing (CIDR). Supernetting provides a way to combine multiple network routes into fewer supernet routes, thus reducing the number of routes in the routing table and in advertisements.

For configuration information for RIP, see “Steps for configuring OSPF and RIP (IPv4 and IPv6)” on page 305.

IPv6 dynamic routing using router discovery

Enabling IPv6 router discovery in z/OS Communications Server requires no additional z/OS Communications Server configuration. All that is needed is at least one IPv6 interface that is defined and started, and at least one adjacent router through that interface that is configured for IPv6 router discovery. If these things exist, then z/OS Communications Server begins receiving router advertisements from the adjacent routers. Depending on the configuration in the adjacent routers, the following types of routes can be learned from the received router advertisements:

- Default route for which the originator of the router advertisement is the next hop
- Indirect routes to prefixes for which the originator of the router advertisement is the next hop
- Direct routes (no next hop) to prefixes that are on the link that is shared by z/OS Communications Server and the originator of the router advertisement

If a stack routing table contains non-replaceable static routes to a destination, any route to that destination that is learned from a router advertisement is not added to the stack routing table.

Multiple routes from router advertisements

Multiple default routes and multiple prefix routes to a single prefix might be learned through router advertisements. If an adjacent router is on a link onto which z/OS Communications Server TCP/IP has multiple IPv6 interfaces, multiple routes are learned for each route in the router advertisement from the adjacent router (one route through each interface onto the link). Default routes and indirect routes to a prefix might be learned from the router advertisements that are originated by multiple adjacent routers.

When multiple direct prefix routes to a single prefix are learned, all of the routes are installed in the stack main route table. The subset of the routes that are installed in a stack policy-based route table is controlled by the DynamicRoutingParms parameter on the RouteTable statement for that route table.

When multiple default routes or multiple indirect prefix routes to a single prefix are learned, the stack uses the following information to determine the subset of the routes that are installed in the stack route tables:

- The reachability of the advertising routers.
- The status (active or inactive) of the associated interfaces.
- The preference values (high, medium, or low) that are advertised for the routes in the router advertisements. This value enables the originator of the router advertisement to indicate whether it is to be preferred over other routers as the next hop for the route.
- For a policy-based route table, the dynamic routing parameters that are defined for the route table by using the DynamicRoutingParms parameter on the RouteTable statement.

The routes with the highest preference value are installed in the route table; however, routes that are learned from routers that are reachable over active interfaces take priority, regardless of preference value.

When multiple default routes or multiple prefix routes to a single prefix are installed in the stack main route table, TCP/IP uses those routes according to the setting of the MULTIPATH parameter on the IPCONFIG6 statement. When multiple default routes or multiple prefix routes to a single prefix are installed in a stack policy-based route table, TCP/IP uses those routes according to the setting of the Multipath6 parameter on the RouteTable statement for that route table.

IPv6 dynamic routing using OMPROUTE

For IPv6, OMPROUTE implements the IPv6 RIP protocol described in RFC 2080 (*RIPng for IPv6*) and the IPv6 OSPF protocol described in RFC 2740 (*OSPF for IPv6*). It provides an alternative to the static TCP/IP gateway definitions. The MVS host running with OMPROUTE becomes an active OSPF or RIP router in an IP network. Either or both of these routing protocols can be used to dynamically maintain the host IPv6 routing table. For example, OMPROUTE can detect when a route is created, is temporarily unavailable, or if a more efficient route exists. If both IPv6 OSPF and IPv6 RIP protocols are used simultaneously, IPv6 OSPF routes will be preferred over IPv6 RIP routes to the same destination.

Tip: You can use the IBM Health Checker for z/OS to check whether the total number of indirect routes in a TCP/IP stack routing table exceeds a maximum threshold. When this threshold is exceeded, OMPROUTE and the TCP/IP stack can experience high CPU consumption from routing changes. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide and IBM Health Checker for z/OS: User's Guide.

IPv6 OSPF protocol

IPv6 OSPF is classified as an Interior Gateway Protocol (IGP). This means that it distributes routing information between routers belonging to a single autonomous system (AS), a group of routers all using a common routing protocol. The IPv6 OSPF protocol is based on link-state or shortest path first (SPF) technology.

IPv6 OSPF is a dynamic routing protocol. It quickly detects topological changes in the AS (such as router interface failures) and calculates new loop-free routes after a period of convergence. This period of convergence is short and involves a minimum of routing traffic as compared to the IPv6 RIP protocol. However, it does generally require more CPU resources on participating routers.

In IPv6 OSPF, sets of networks can be placed together in groups called areas. The topology of an area is hidden from the rest of the AS. This method of hiding information enables a significant reduction in routing traffic. Also, routing within the area is determined only by the area's own topology, lending the area protection from bad routing data.

Each IPv6 OSPF router maintains a database describing the autonomous system's topology. Each individual piece of this database is a particular router's local state (for example, the router's usable interfaces and reachable neighbors). The router distributes its local state information by flooding. The routing information in an area is summarized and advertised into adjacent areas, allowing for the generation of interarea routes. This summarization and advertising of routing information between areas is the responsibility of the routers that are on the border between areas (called area border routers).

To generate routes, all routers run the exact same algorithm, in parallel. From the topological database, each router constructs a tree of shortest paths with itself as root. This shortest-path tree gives the route to each destination in the autonomous system. Externally derived routing information (for example, routes learned from the IPv6 RIP protocol) appears on the tree as leaves. When multiple equal-cost routes to a destination exist, the routes (up to 16) are added to the TCP/IP stack's route table. The TCP/IP stack uses these equal-cost routes according to the multipath setting configured for the route table. “Multiple equal-cost routes” on page 295 provides additional information about the multipath setting configured for a route table and the use of multiple equal-cost routes.

Externally derived routing data is passed transparently throughout the autonomous system. This externally derived data is kept separate from the IPv6 OSPF protocol's link state data. Each external route can also be tagged by the advertising router, but not by OMPROUTE, enabling the passing of additional information between routers on the boundaries of the autonomous system. OMPROUTE does pass tags created by others. For information on configuring IPv6 OSPF, see “Steps for configuring OSPF and RIP (IPv4 and IPv6)” on page 305.

IPv6 RIP protocol

IPv6 RIP is an Interior Gateway Protocol (IGP) designed to manage a relatively small network. IPv6 RIP is based on the Bellman-Ford or the distance-vector algorithm. IPv6 RIP has limitations and is not suited for every TCP/IP environment. Before using the IPv6 RIP function in OMPROUTE, read RFC 2080 to decide if RIP can be used to manage the IPv6 routing tables of your network. For more information about RFC 2080, see Appendix G, “Related protocol specifications,” on page 1409.

IPv6 RIP uses the number of hops, or hop count, to determine the best possible route to a host or network. The term hop count is also referred to as the metric. In IPv6 RIP, a hop count of 16 means infinity, or that the destination cannot be reached. This limits the longest path in the network that can be managed by IPv6 RIP to 15 gateways.

A IPv6 RIP router broadcasts routing information to its directly connected networks every 30 seconds. It receives updates from neighboring IPv6 RIP routers every 30 seconds and uses the information contained

in these updates to maintain the IPv6 routing table. If an IPv6 RIP update has not been received from a neighboring router in 180 seconds, an IPv6 RIP router assumes that the neighboring router is down, sets all IPv6 RIP routes through that router to a metric of 16 (infinity), and stops using those routes when routing IP packets. If an update has still not been received from the neighboring router after another 120 seconds, the IPv6 RIP router deletes from the IPv6 routing table all of the IPv6 RIP routes through that neighboring router.

Next hop IP addresses, whenever applicable, are included in IPv6 RIP updates to eliminate packets being routed through extra hops in the network. IPv6 multicast address FF02::9, reserved for IPv6 RIP packets, is used to reduce unnecessary load on hosts that are not listening for IPv6 RIP messages.

For configuration information for IPv6 RIP, see “Steps for configuring OSPF and RIP (IPv4 and IPv6)” on page 305.

OMPROUTE configuration

This topic includes items to consider when configuring OMPROUTE, and the steps to use to configure OMPROUTE.

Run-time environment

OMPROUTE is a z/OS UNIX application, and it requires a z/OS UNIX file system to operate. It can be started from an MVS started procedure, from the z/OS shell, or from AUTOLOG (see step “2” on page 297 for restrictions on using AUTOLOG to start OMPROUTE). OMPROUTE must be started by a RACF-authorized user ID, and it must be in an APF authorized library.

OMPROUTE uses the MVS operator's console, syslogd, CTRACE, and STDOUT for its logging and tracing. The MVS operator's console and syslogd are used for major events such as initialization, termination, and error conditions. CTRACE is used for tracing the receipt and transmission of OSPF/RIP packets and communications between OMPROUTE and the TCP/IP stack. In addition, OMPROUTE can be configured to use CTRACE for detailed tracing and debugging. STDOUT is used for detailed tracing and debugging.

When syslogd is active and /dev/console is defined in the syslog.conf file, OMPROUTE logging messages are directed to syslogd and to the MVS console. If syslogd is active and /dev/console is not defined, OMPROUTE logging messages are directed to syslogd only.

When syslogd is inactive, OMPROUTE logging messages are directed to the OMPROUTE JES joblog and to the MVS console.

If OMPROUTE tracing or debugging is enabled, the output can be directed to STDOUT, the JES joblog, or OMPROUTE CTRACE. Additionally, the logging messages (that is, interface initialization messages) can be found in syslogd, provided it is active at the time of tracing or debugging. Syslogd directs urgent OMPROUTE messages to the MVS console.

Tip: If you enable OMPROUTE tracing without syslogd active, large amounts of data can be written to the console.

OMPROUTE uses a standard message catalog. The message catalog must be in the z/OS UNIX file system. The directory location for the message catalog path is set by the environment variables NLSPATH and LANG.

Configuration of OMPROUTE is accomplished using an OMPROUTE configuration file. For details on the statements in the OMPROUTE configuration file, see z/OS Communications Server: IP Configuration Reference.

Display of OMPROUTE information is performed using the DISPLAY or MODIFY command. Modification of OMPROUTE information is performed using the MODIFY command. For details on OMPROUTE's DISPLAY and MODIFY commands, see z/OS Communications Server: IP System Administrator's Commands.

Language Environment run-time considerations

When starting OMPROUTE from a started or cataloged procedure, it is usually recommended that OMPROUTE be started directly from the SEZALOAD data set using PGM=OMPROUTE. However, there is a situation where it might be desirable to start OMPROUTE using BPXBATCH.

When OMPROUTE is started using PGM=OMPROUTE, the STDENV DD card, if used, is passed directly to the OMPROUTE program. The Language Environment does not get access to the STDENV environment variables. As a result, any Language Environment run-time options set in the STDENV DD data set using the _CEE_RUNOPTS= environment variable are ignored. In this case, Language Environment run-time options must be passed on the PARM= parameter and the options must be specified before any OMPROUTE options. However, the PARM= statement allows a maximum of 100 characters. If the wanted Language Environment run-time options plus OMPROUTE parameters exceeds 100 characters, consideration should be given to using BPXBATCH to start OMPROUTE. When PGM=BPXBATCH is used, the Language Environment environment variable _CEE_RUNOPTS can be included on the STDENV DD card to specify run-time options in excess of 100 characters long.

OMPROUTE tuning considerations

It might be desirable to tune OMPROUTE's storage usage when running in complex network environments or when running with traces enabled for long periods of time. To determine if tuning is necessary, use the Language Environment run-time options RPTSTG and RPTOPTS. For information on these options and the reports they generate, see z/OS Language Environment documentation.

Multiple TCP/IP stacks

A one-to-one relationship exists between an instance of OMPROUTE and a stack. OSPF, RIP, IPv6 OSPF, and IPv6 RIP support on multiple stacks requires multiple instances of OMPROUTE.

TCP/IP stack routing table management

OMPROUTE's job is limited to the management of the TCP/IP stack routing table. OMPROUTE is not involved in the actual routing decisions made by the TCP/IP stack when routing a packet to its destination.

If IPv4 interfaces are defined in the OMPROUTE configuration file as OSPF or RIP interfaces, all IPv4 dynamic routes are deleted from the stack's IPv4 routing table upon initialization of OMPROUTE. If IPv6 interfaces are defined in the OMPROUTE configuration file as OSPF or RIP interfaces, all IPv6 dynamic routes, excluding those dynamic routes learned through IPv6 router discovery, are deleted from the stack's IPv6 routing table upon initialization of OMPROUTE. OMPROUTE then repopulates the stack routing tables that it cleared, using information learned through the routing protocols.

IPv4 Internet Control Message Protocol (ICMP) redirects are ignored when OMPROUTE is active and there are IPv4 interfaces configured to OMPROUTE as RIP or OSPF interfaces. IPv6 ICMP redirects are ignored when OMPROUTE is active and there are IPv6 interfaces configured to OMPROUTE as RIP or OSPF interfaces.

OMPROUTE does not make use of the BSDROUTINGPARMS statement. Instead, the IPv4 Maximum Transmission Unit (MTU), subnet mask, and destination address parameters are configured using the OSPF_INTERFACE, RIP_INTERFACE, and INTERFACE statements in the OMPROUTE configuration file. The MTU for IPv6 interfaces is learned from the TCP/IP stack and therefore is not a parameter on the IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, and IPV6_INTERFACE statements.

Result: The Netstat DEvlinks/-d report displays these parameters under the heading `Routing Parameters`, even though the BSDROUTINGPARMS statement is not defined.

Using RIP, IPv6 RIP, OSPF, and IPv6 OSPF with OMPROUTE

When OMPROUTE is initialized, it uses the OMPROUTE configuration file to determine which routing protocols will be enabled. If at least one OSPF interface is configured, the OSPF protocol is enabled. If at least one RIP interface is configured, RIP is enabled. If at least one IPv6 RIP interface is configured, IPv6 RIP is enabled. If at least one IPv6 OSPF interface is configured, IPv6 OSPF is enabled. If OMPROUTE

is started with no interfaces defined for a particular protocol, that protocol is disabled until one of the following events occur:

- OMPROUTE is stopped and restarted with a configuration file containing at least one interface of the specific type.
- OMPROUTE is dynamically reconfigured using the MODIFY command with a configuration file containing at least one interface of the specific type.

When OMPROUTE is configured for both the OSPF and RIP protocols (either IPv4 or IPv6), routes that are learned through the OSPF protocol take precedence over routes learned through the RIP protocol.

The OSPF and RIP protocols are communicated over IPv4 interfaces that are defined with the OSPF_INTERFACE and RIP_INTERFACE configuration statements, respectively. An IPv4 interface involved in the communication of neither the RIP nor the OSPF protocol should be configured to OMPROUTE with the INTERFACE configuration statement, unless Ignore_Undefined_Interfaces=YES is coded on the Global_Options configuration statement. OMPROUTE supports a maximum of 255 real, physical, IPv4 interfaces (that is, interfaces on which data can actually be sent and received). There is no theoretical limit on how many VIPAs can be configured, though there are practical limits imposed by network design. For special VIPA considerations, see step “5” on page 309.

The IPv6 OSPF and IPv6 RIP protocols are communicated over IPv6 interfaces that are defined with the IPV6_OSPF_INTERFACE and IPV6_RIP_INTERFACE configuration statements, respectively. An IPv6 interface not involved in the communication of the IPv6 OSPF or IPv6 RIP protocol can be configured to OMPROUTE with the IPV6_INTERFACE configuration statement. If default values are acceptable and you do not need to define additional prefixes to an IPv6 interface not involved in the communication of the IPv6 OSPF or IPv6 RIP protocol, it is not necessary to configure the interface to OMPROUTE at all. OMPROUTE will learn about the interface and its MTU value from the stack and use default values for other parameters. This is different from IPv4, where all interfaces should be configured to prevent OMPROUTE from using default values for MTU size and subnet mask, unless Global_Options is coded with Ignore_Undefined_Interfaces=YES.

OMPROUTE allows for the generation of multiple, equal-cost routes to a destination. For OSPF and IPv6 OSPF, up to 16 multiple equal-cost routes are allowed. For RIP and IPv6 RIP, multiple equal-cost routes are supported only to directly connected destinations over redundant interfaces.

Virtual IP addresses

OMPROUTE is enhanced with virtual IP addressing (VIPA) to handle network interface failures by switching to alternate paths. The VIPA routes are included in the OSPF, RIP, IPv6 OSPF, and IPv6 RIP advertisements to adjacent routers. Adjacent routers learn about VIPA routes from the advertisements and can use them to reach the destinations at the MVS host.

Service policy

If service policy is going to be used to restrict access to neighbors on point-to-multipoint interfaces (for example MPCPTP interfaces including XCF and IUTSAMEH connections) for temporary intervals, those neighbors must be explicitly defined on the OSPF_INTERFACE or RIP_INTERFACE statement. Otherwise, OMPROUTE might not be able to communicate with those neighbors when the access restriction expires.

Multiple equal-cost routes

When the TCP/IP main route table is used to route traffic, multiple routes exist in the TCP/IP main route table for a destination, and the IPCONFIG MULTIPATH or IPCONFIG6 MULTIPATH statement is specified in PROFILE.TCPIP, outbound traffic for that destination is distributed across all of the routes. The same is true when a policy-based route table is used to route traffic, multiple routes exist in the policy-based route table for a destination, and the Multipath parameter on the RouteTable policy statement indicates that multipath support should be provided for the policy-based route table. This traffic distribution is done on either a packet-basis or connection-basis, depending on the type of multipath support configured. For information on configuring different types of multipath support using the IPCONFIG and IPCONFIG6

statements, and the RouteTable policy statement, see z/OS Communications Server: IP Configuration Reference.

When OMPROUTE is being used to provide dynamic routing for a TCP/IP stack, multiple routes to the same destination can be dynamically added to a TCP/IP stack route table, based upon the routing information learned from other routers. These multiple routes will be added when the route calculation for each has resulted in the same route cost value. No more than 16 equal-cost routes will be added for each destination. For RIP and IPv6 RIP, multiple equal-cost routes will be added only to directly-connected destinations over redundant interfaces. The RIP and IPv6 RIP protocols will generate no more than one indirect route to a destination.

Table 18. Multipath route limitations

Multipath route type	Static (BEGINROUTES and policy-defined)	OMPROUTE (OSPF and IPv6 OSPF)	OMPROUTE (RIP and IPv6 RIP)
Direct host	Yes (no limit)	Yes (up to 16)	No
Indirect host	Yes (no limit)	Yes (up to 16)	No
Direct network	Yes (no limit)	Yes (up to 16)	Yes (up to 16 for redundant interfaces)
Indirect network	Yes (no limit)	Yes (up to 16)	No
Default (indirect)	Yes (no limit)	Yes (up to 16)	No

Guidelines:

- If more equal-cost multipath routes are needed in the main IPv4 route table than can be provided by an IPv4 dynamic routing protocol, you can configure static routes using the BEGINROUTES statement.
- If more equal-cost multipath routes are needed in the main IPv6 route table than can be provided by an IPv6 dynamic routing protocol, you can configure static routes using the BEGINROUTES statement.
- If more equal-cost multipath routes are needed in a policy-based route table than can be provided by a dynamic routing protocol, you can configure static routes using the RouteTable policy statement.

Sysplex autonomics

If you enable sysplex autonomics, it is important that you ensure that the WLM policy for the OMPROUTE address space receives sufficient resources in relationship to other work that is being managed on the system. Under high load conditions, if not properly classified it is possible that OMPROUTE can trigger an autonomic response from the TCP/IP stack it has an affinity with, resulting in the TCP/IP address space removing itself from the sysplex group. Place the TCP/IP and OMPROUTE address spaces in the SYSSTC service classification. Classification in another service class leaves the system vulnerable to a sysplex distributor outage. For more information about sysplex autonomics, see "Sysplex problem detection and recovery" on page 452.

Steps for configuring OMPROUTE

OMPROUTE is a multiprotocol routing daemon capable of handling both OSPF and RIP interfaces concurrently.

Before you begin

You can configure OMPROUTE by using the following general steps. Details follow the general steps.

1. Create the OMPROUTE configuration file.
2. Reserve the ports, and ensure loopback availability.
3. Update the resolver configuration file.
4. Update the OMPROUTE cataloged procedure.

5. Specify the RIP UDP port numbers in the SERVICES file or data set (if using the RIP or IPv6 RIP protocol).
6. RACF authorize user IDs for starting OMPROUTE.
7. Start syslogd.
8. Update the OMPROUTE environment variables (optional).
9. Create static routes (optional).
10. Configure OSPF authentication (optional, if using the IPv4 OSPF protocol).

If policy-based routing is used on the TCP/IP stack and the routing policy is configured with dynamic routing parameters, you do not need to provide additional configuration to OMPROUTE for dynamic routing support to be provided for the policy-based route tables. The dynamic routing parameters specified in the routing policy are provided to OMPROUTE by the TCP/IP stack to control the scope of dynamic routes computed by OMPROUTE. For a description of this function, see "Policy-based routing" on page 349.

If policy-based routing is used on the TCP/IP stack and a route table is configured to the Policy Agent with no dynamic routing parameters, OMPROUTE has no knowledge of that route table. For example, the route table does not appear in the display of OMPROUTE route tables.

Procedure

Perform the following steps to configure OMPROUTE:

1. Create the OMPROUTE configuration file.

   The OMPROUTE configuration file provides information about the host's routing capabilities and TCP/IP interfaces. For more detail about the contents of this file, see "Steps for configuring OSPF and RIP (IPv4 and IPv6)" on page 305. The following search order is used by OMPROUTE to locate the configuration data set or file:

   a. The MVS data set or z/OS UNIX file that is specified on the OMPCFG DD statement in the OMPROUTE started procedure.
   b. If the environment variable, OMPROUTE_FILE, is defined, OMPROUTE uses the value as the name of an MVS data set or z/OS UNIX file to access the configuration data. The syntax for an MVS data set name is //'*mvs.dataset.name*'. The syntax for a z/OS UNIX file name is */dir/subdir/file.name*.
   c. /etc/omproute.conf
   d. *hlq*.ETC.OMPROUTE.CONF

   Tip: If you are configuring a complex environment, you can use the following OMPROUTE configuration file features:

   - Group related statements into separate files and use the INCLUDE statement to include them in your OMPROUTE configuration.
   - Use MVS system symbols (such as &SYSCLONE, &SYSNAME, and &SYSPLEX). This feature reduces the number of OMPROUTE configuration files that must be maintained in a multisystem environment.

   A sample configuration file is provided in SEZAINST(EZAORCFG). The configuration files for TCPCS4, TCPCS6, TCPCS7, and TCPCS64 in the sample network are shown in "Sample OMPROUTE configuration files" on page 347. For a description of the syntax rules for the OMPROUTE configuration file and details on each of the configuration statements, see z/OS Communications Server: IP Configuration Reference.
2. Reserve the ports.

   - RIP protocol

     If the RIP protocol of OMPROUTE is going to be used, UDP port 520 should be reserved for OMPROUTE. If the IPv6 RIP protocol of OMPROUTE is going to be used, UDP port 521 should be reserved for OMPROUTE. This is done by adding the name of the member containing the OMPROUTE cataloged procedure to the PORT statement in PROFILE.TCPIP:

```
PORT
    520 UDP OMPROUTE
    521 UDP OMPROUTE
```

If you want to be able to start OMPROUTE from the z/OS shell, use the special name OMVS as follows:

```
 PORT
    520 UDP OMVS
    521 UDP OMVS
```

- Autolog considerations for OMPROUTE when using the OSPF protocol

  If a procedure in the AUTOLOG list also has a PORT statement reserving a TCP or UDP port but does not have a listening connection on that port, TCP/IP periodically attempts to cancel that procedure and start it again.

  Therefore, if OMPROUTE is being started with AUTOLOG and only the OSPF or IPv6 OSPF protocol is being used (no RIP or IPv6 RIP protocol, so no listening connection on the RIP or IPv6 RIP UDP port), it is important to take one of the following actions:

  – Ensure that the RIP UDP port (520) and the IPv6 RIP UDP port (521) are not reserved by the PORT statement in the PROFILE.TCPIP.
  – Add the NOAUTOLOG parameter to the PORT statement in the PROFILE.TCPIP. For example,

    ```
    PORT
     520 UDP OMPROUTE NOAUTOLOG
     521 UDP OMPROUTE NOAUTOLOG
    ```

  Tip: When using only the OSPF or IPv6 OSPF protocol, the auto-start feature of AUTOLOG can be used as described. However, the monitoring and auto-restart features of AUTOLOG are unavailable due to AUTOLOG's dependence on listening to a TCP or UDP connection, which does not exist with OSPF and IPv6 OSPF.

  If you fail to take one of these actions, OMPROUTE will be periodically canceled and restarted by TCP/IP.

3. Update the resolver configuration file.

   The resolver configuration file contains keywords (DATASETPREFIX and TCPIPjobname) used by OMPROUTE. The value assigned to DATASETPREFIX will determine the high-level qualifier (*hlq*). The *hlq* is used in the search order for the OMPROUTE configuration file. If no DATASETPREFIX keyword is found, a default of TCPIP is used. The value assigned to TCPIPjobname will be used as the name of the TCP/IP stack with which OMPROUTE establishes a connection.

   For a description of the search order used by the resolver to locate the resolver configuration file, see “Resolver configuration files” on page 771.

4. Update the OMPROUTE cataloged procedure.

   If OMPROUTE is to be started by a procedure, create the cataloged procedure by copying the sample in SEZAINST(OMPROUTE) to your system or recognized PROCLIB. Specify OMPROUTE parameters and change the data set names to suit your local configuration.

```
//*
//* TCP/IP for MVS
//* SMP/E Distribution Name: EZBORPRC
//* //*        5650-ZOS Copyright IBM Corp. 1998, 2015
//*        Licensed Materials - Property of IBM
//*        This product contains "Restricted Materials of IBM"
//*        All rights reserved.
//*        US Government Users Restricted Rights -
//*        Use, duplication or disclosure restricted by
//*        GSA ADP Schedule Contract with IBM Corp.
//*        See IBM Copyright Instructions.
//*
//OMPROUTE PROC
//OMPROUTE EXEC PGM=OMPROUTE,REGION=10M,TIME=NOLIMIT,
```

```
// PARM=('ENVAR("_CEE_ENVFILE_S=DD:STDENV")/')
//*
//* Example of start parameters to OMPROUTE:
//*
//* PARM=('ENVAR("_CEE_ENVFILE_S=DD:STDENV")/-t1 -6t1')
//*
//*        Provide environment variables to run with the desired
//*        stack and configuration.  As an example, the file
//*        specified by STDENV could have these lines in it:
//*
//*           RESOLVER_CONFIG=//'SYS1.TCPPARMS(TCPDATA2)'
//*           OMPROUTE_FILE=/u/usernnn/config.tcpcs2
//*           OMPROUTE_DEBUG_FILE=/tmp/logs/omproute.debug
//*           OMPROUTE_IPV6_DEBUG_FILE=/tmp/logs/omprout6.debug
//*           OMPROUTE_DEBUG_FILE_CONTROL=1000,5
//*
//*        For information on the above environment variables,
//*        refer to the IP CONFIGURATION GUIDE.
//*
//*        When using _CEE_ENVFILE with an MVS data set, the data   //*        set must
be allocated with RECFM=V.  To use a RECFM=F    //*        data set, _CEE_ENVFILE_S
should be used to prevent the   //*        environment variable values from being padded
with        //*        blanks.     //*
//*        If you want to include comments in the data set or
//*        z/OS UNIX file, specify the _CEE_ENVFILE_COMMENT
//*        environment variable as the first environment variable
//*        in the data set or file.  The value specified for the
//*        _CEE_ENVFILE_COMMENT variable is the comment character.
//*        For example, if you want to use the semicolon sign, ;,
//*        as the comment character, specify this as the first
//*        statement:
//*           _CEE_ENVFILE_COMMENT=;
//*
//STDENV   DD PATH='/u/usernnn/envcs2',
//         PATHOPTS=(ORDONLY)
//*
//*        The stdout stream may be redirected to a HFS file as
//*        shown below.
//*        The PATHOPTS OTRUNC option will clear the stdout file
//*        every time OMPROUTE is started.  If you want to retain
//*        previous stdout information, change it to OAPPEND.
//*
//SYSPRINT DD SYSOUT=*
//*SYSPRINT DD PATH='/tmp/omproute.stdout',
//*            PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//*            PATHMODE=(SIRUSR,SIWUSR,SIRGRP,SIWGRP)
//*
//*        The stderr stream may be redirected to a HFS file as
//*        shown below.
//*        The PATHOPTS OTRUNC option will clear the stderr file
//*        every time OMPROUTE is started.  If you want to retain
//*        previous stderr information, change it to OAPPEND.
//*
//SYSOUT   DD SYSOUT=*
//*SYSOUT   DD PATH='/tmp/omproute.stderr',
//*            PATHOPTS=(OWRONLY,OCREAT,OTRUNC),
//*            PATHMODE=(SIRUSR,SIWUSR,SIRGRP,SIWGRP)
//*
//CEEDUMP  DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
```

5. Specify the RIP UDP port numbers in the SERVICES file or data set (if using the RIP or IPv6 RIP protocol).

   The services file contains the relationship between services and port numbers as described in z/OS Communications Server: IP Configuration Reference. The portion of the services file relevant to OMPROUTE is:

   ```
   route           520/udp         router omproute
   route           521/udp         ipv6rip ripng
   ```

   The file must exist for the RIP and IPv6 RIP protocols of OMPROUTE to operate.

   For a description of the search order used to locate the services file, see “Configuration files for TCP/IP applications” on page 25.

6. RACF authorize user IDs for starting OMPROUTE.

To reduce risk of an unauthorized user starting OMPROUTE and affecting the contents of the routing table, users who start OMPROUTE must be RACF-authorized to the entity MVS.ROUTEMGR.OMPROUTE. To RACF-authorize, the following commands must be entered from a RACF user ID, substituting the authorized user ID on the ID(userid) parameter. The commands in this example are taken from SEZAINST(EZARACF).

```
RDEFINE OPERCMDS (MVS.ROUTEMGR.OMPROUTE) UACC(NONE)
PERMIT MVS.ROUTEMGR.OMPROUTE ACCESS(CONTROL) CLASS(OPERCMDS) ID(userid)
SETROPTS RACLIST(OPERCMDS) REFRESH
```

Rule: The user ID specified for OMPROUTE must have read/write permission to the directories where OMPROUTE reads and writes files. To determine the files that OMPROUTE reads and writes, see OMPROUTE environment variables in z/OS Communications Server: IP Configuration Reference.

7. Start syslogd.

   To write only the urgent OMPROUTE messages to the z/OS console, syslogd should be running while OMPROUTE is running. Syslogd sends the non-urgent messages to the z/OS UNIX file system message log.

8. Update the OMPROUTE environment variables (optional).

   The following environment variables are used by OMPROUTE and can be tailored to a particular installation:

   RESOLVER_CONFIG
   The RESOLVER_CONFIG variable is used by OMPROUTE to locate the resolver configuration file. For more information on OMPROUTE's use of the resolver configuration file, see step “3” on page 298. For more information about the RESOLVER_CONFIG environment variable, see “z/OS XL C/C++ environment variables for configuration files” on page 775.

   OMPROUTE_FILE
   The OMPROUTE_FILE variable is used by OMPROUTE in the search order for the OMPROUTE configuration file. For details on the search order used for locating this configuration file, see step “1” on page 297.

   OMPROUTE_DEBUG_FILE
   The OMPROUTE_DEBUG_FILE variable is used by OMPROUTE to override the debug output destination for IPv4 dynamic routing protocols and for processing common to both IPv4 and IPv6 routing protocols.

   Tip: When the OMPROUTE CTRACE with option DEBUGTRC (or option ALL) is active, debug output is written to CTRACE and not to the destination specified by this variable.

   For more information on using this environment variable, see “OMPROUTE parameters” on page 302.

   OMPROUTE_DEBUG_FILE_CONTROL
   The OMPROUTE_DEBUG_FILE_CONTROL variable is used by OMPROUTE to control the size and quantity of trace files created when the OMPROUTE_DEBUG_FILE or OMPROUTE_IPV6_DEBUG_FILE variable is specified. The syntax of this variable is:

   ```
   OMPROUTE_DEBUG_FILE_CONTROL=<size of file>,<num of files>
   ```

   The default values for <size of file> and <num of files> are 200 (KB) and 5 respectively. In general, these values are sufficient for most installations.

   If OMPROUTE_DEBUG_FILE and OMPROUTE_IPV6_DEBUG_FILE are both specified with different output destinations, the values specified on the OMPROUTE_DEBUG_FILE_CONTROL variable will apply to both the IPv4 debug files and the IPv6 debug files. The total number of files created in this environment would be <num of files> multiplied by 2.

   OMPROUTE_IPV6_DEBUG_FILE
   The OMPROUTE_IPV6_DEBUG_FILE variable is used by OMPROUTE to override the debug output destination for IPv6 dynamic routing protocols.

Tip: When the OMPROUTE CTRACE with option DEBUGTRC (or option ALL) is active, debug output is written to CTRACE and not to the destination specified by this variable.

For more information on using this environment variable, see "OMPROUTE parameters" on page 302.

OMPROUTE_CTRACE_MEMBER
The OMPROUTE_CTRACE_MEMBER variable is used by OMPROUTE to specify the name of the parmlib member containing CTRACE default settings. If not specified, the default value is CTIORA00. Use this environment variable to set different CTRACE options and buffer sizes for multiple OMPROUTE instances.

9. Create static routes (optional).

   To create IPv4 or IPv6 static routes, use the BEGINROUTES statement in PROFILE.TCPIP. For information about the syntax of the BEGINROUTES statement, see z/OS Communications Server: IP Configuration Reference.

   During initialization, OMPROUTE learns of static routes by reading the internal routing table set up by TCP/IP. If static routes are changed during execution by VARY TCPIP,,OBEYFILE command processing, OMPROUTE is dynamically notified of the changes by TCP/IP. OMPROUTE will advertise active static routes to other routers if allowed by configuration (for example, the IMPORT_STATIC_ROUTES parameter of the AS_BOUNDARY_ROUTING or IPV6_AS_BOUNDARY_ROUTING configuration statements).

   Static routes can be defined as replaceable or nonreplaceable, with nonreplaceable being the default.

   A nonreplaceable static route cannot be replaced or modified by OMPROUTE, even if a better dynamic route can be learned and even if the static route is not actually available (but a static route that is not available will not be advertised by OMPROUTE). Because of this, the use of nonreplaceable static routes with OMPROUTE is not recommended unless it is to provide routing over an interface over which no routing protocol is being communicated.

   A replaceable static route will be replaced by OMPROUTE if it dynamically learns a route to the destination. Any dynamically learned route will be considered more desirable than a replaceable static route. A replaceable static route should be considered as a last resort route, to be used by TCP/IP when no dynamic route to a destination can be found.

   For detailed information, see "Use of static routing with OMPROUTE" on page 357.

10. Configure OSPF authentication (optional, if using the IPv4 OSPF protocol).

    OMPROUTE supports defining the IPv4 OSPF authentication type by area or by interface. By default, all interfaces attached to an area use the type of authentication defined for that area on the AREA configuration statement, unless overridden on the OSPF_INTERFACE configuration statement. The values of authentication keys must be defined on OSPF_INTERFACE statements in any case. All routers which could become neighbors of each other must use the same authentication type and key, or OSPF communication between the routers will not be possible.

    Virtual links behave similarly to interfaces for authentication purposes. By default, a virtual link uses the same type of authentication that is specified for the backbone area unless overridden on the VIRTUAL_LINK configuration statement. When the authentication type is not NONE, the value of the authentication key must be specified on the VIRTUAL_LINK configuration statement. There is no requirement for a virtual link to have the same authentication key value as its underlying real interface.

    OSPF authentication does not protect the contents of an OSPF packet. These packets are not encrypted. However, it does provide verification that the packet is genuine.

    Tip: Unlike IPv4 OSPF, IPv6 OSPF does not provide its own, protocol-layer authentication. It relies instead on authentication provided by IPv6 IPSec.

    There are two methods of IPv4 OSPF authentication, password and MD5 cryptographic.

    Password authentication is very basic: an 8-byte password is appended to all OSPF packets and sent in the clear with the rest of the packet. If the sent password matches that defined by the packet receiver, the packet is accepted.

MD5 authentication is more sophisticated. The combination of the OSPF packet and the MD5 key is summarized into a 16-byte message digest, which is appended to the packet and sent. The keys are never sent, only the message digests. The receiver then attempts to re-create the message digest from the combination of its defined key and the OSPF packet. If the digest is successfully re-created, the packet is accepted; otherwise it is rejected. MD5 authentication also contains a monotonic increasing counter to protect against replay attacks.

If MD5 cryptographic authentication is being used, a 16-byte MD5 key must be defined on the OSPF_INTERFACE configuration statement. This key is defined as a hexadecimal string and may be obtained in several ways. One method for obtaining MD5 keys is provided in the pwtokey utility, which converts a password into an MD5 key. This UNIX System Services utility implements the algorithm defined in RFC 3414, *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)*. Because OSPF does not support localization of keys, it is necessary only to provide a password to this utility to generate a single, 16-byte key. If multiple sites have this utility, MD5 keys can easily be generated from passwords, which are easier to remember and communicate than 16-byte hexadecimal strings.

Starting and controlling OMPROUTE

After the necessary RACF authorization has been defined (see step "6" on page 299), OMPROUTE can be started from an MVS procedure, from the z/OS shell, or from AUTOLOG.

- You can start OMPROUTE from the MVS operators console by starting the OMPROUTE start procedure. A sample start procedure is provided with the product in SEZAINST(OMPROUTE).
- You can start OMPROUTE from the z/OS shell by starting OMVS and then issuing the OMPROUTE command and, optionally, any parameters. For information on parameters, see "OMPROUTE parameters" on page 302.
- You can use the AUTOLOG statement to start OMPROUTE automatically during TCP/IP initialization. Insert the name of the OMPROUTE start procedure in the AUTOLOG statement of the PROFILE.TCPIP data set.

```
AUTOLOG
  OMPROUTE
ENDAUTOLOG
```

Note: For special considerations when using AUTOLOG to start OMPROUTE with the OSPF or IPv6 OSPF protocol, see step "2" on page 297.

In a Common INET environment, OMPROUTE will attempt to connect to a stack whose name is determined by the TCPIPjobname keyword from the resolver configuration data set or file. In configurations with multiple stacks, a copy of OMPROUTE must be started for each stack that requires OMPROUTE services. To associate OMPROUTE with a particular stack, use the environment variable RESOLVER_CONFIG to point to the data set or file that defines the unique TCPIPjobname.

When running from an MVS procedure, the environment variables can be set by using the STDENV DD statement in the OMPROUTE procedure. For information concerning the environment variables used by OMPROUTE, see step "8" on page 300.

OMPROUTE parameters

OMPROUTE accepts five command line parameters, which govern tracing and debug information. OMPROUTE's trace and debug information is written to stdout with three exceptions:

- When OMPROUTE is started with no tracing, and then a MODIFY command is issued to enable tracing. In this case, by default, the output destination is the file omproute_debug in the current temporary directory (the default is /tmp).
- When the debug output destination has been overridden with the use of an environment variable (OMPROUTE_DEBUG_FILE or OMPROUTE_IPV6_DEBUG_FILE).

Rule: OMPROUTE_DEBUG_FILE and OMPROUTE_IPV6_DEBUG_FILE can specify only a z/OS UNIX file. To have debug information go into an MVS data set, instead of coding these environment variables, use the SYSPRINT DD card to redirect stdout to the wanted MVS data set.

- When the OMPROUTE CTRACE with option DEBUGTRC (or option ALL) is enabled. In this case, the output is sent to the CTRACE facility.

If OMPROUTE is to be started from an MVS procedure, add your parameters to PARM=() in the OMPROUTE cataloged procedure. For example:

```
//* PARM=('POSIX(ON)',
//*       'ENVAR("_CEE_ENVFILE=DD:STDENV")/-t1 -6t1')
//*
```

If OMPROUTE is to be started from a z/OS shell command line, enter the parameters on the command line.

For either method of starting OMPROUTE, parameters can be specified in mixed case.

Note: The -t*n*, -d*n*, -6t*n*, -6d*n*, and -s*n* parameters affect OMPROUTE performance. If you use these parameters, you might need to increase the Dead_Router_Interval value on OSPF interfaces to prevent neighbor adjacencies from collapsing. If this becomes necessary, you will also need to modify other routers because the Dead_Router_Interval value must be the same for all routers attached to a common network.

The -tn and -6tn command line parameters

The -t*n* parameter specifies the external tracing level for OMPROUTE initialization and IPv4 routing protocols, where *n* is a supported trace level. The -6t*n* parameter specifies the external tracing level for IPv6 routing protocols, where *n* is a supported trace level. The -t*n* and -6t*n* traces are intended for customers, testers, service, or developers, and provide information on the operation of the routing application. These options can be used for many purposes, such as debugging a configuration, education on the operation of the routing application, verification of test cases, and so on. The following levels are supported:

1
Informational messages

2
Formatted packet trace

These option levels are cumulative—level 2 includes level 1. For example, -t2 provides formatted packet trace and informational messages.

The -dn, -6dn, and -sn command line parameters

These options specify the internal debugging levels. They are intended for service and provide internal debugging information needed for debugging problems. Use of these parameters can significantly impact performance and are not recommended unless needed to debug a problem. For more information about the use of these parameters, see z/OS Communications Server: IP Diagnosis Guide.

Controlling OMPROUTE

You can control OMPROUTE from the operator's console using the MODIFY command. The syntax of the MODIFY command can be found in z/OS Communications Server: IP System Administrator's Commands. MODIFY commands are available to perform the following functions:

- “Stopping OMPROUTE” on page 304
- “Rereading the configuration file” on page 304
- “Enabling or disabling the OMPROUTE subagent” on page 304
- “Changing the cost of OSPF links” on page 305
- “Controlling OMPROUTE tracing and debugging” on page 305

Stopping OMPROUTE

OMPROUTE can be stopped in several ways:

- From MVS, issue STOP <procname> or MODIFY <procname>,KILL.

  If OMPROUTE was started from a cataloged procedure, *procname* is the member name of that procedure. If OMPROUTE was started from the z/OS shell, *procname* is based on the *userid*. If the *userid* is 8 characters long, the *procname* is the *userid*. If the *userid* is less than 8 characters long, the *procname* is useridX, where X is the sequence number set by the system. To determine the sequence number, from the SDSF LOG window on TSO, issue /D OMVS,U=*userid*. This will show the programs running under this user ID. The *procname* can also be set using the environment variable _BPX_JOBNAME and then starting OMPROUTE in the shell background.
- From a z/OS shell superuser ID, issue the kill command to the process ID (PID) associated with OMPROUTE. To determine the PID, use one of the following methods:
  - From the MVS console, issue D OMVS,U=userid, or issue /D OMVS,U=userid at the SDSF LOG window on TSO (where userid is the user ID that started omproute from the shell).
  - Issue the **ps -ef** command from the z/OS shell.
  - Record the PID when you start OMPROUTE.

For information on the environment variable _BPX_JOBNAME, see z/OS UNIX System Services Planning. For information on the D OMVS,U=userid command, see z/OS MVS System Commands.

Tip: When OMPROUTE is taken down, it should be kept down for at least 3 times the largest dead router interval of any IPv4 OSPF interfaces using MD5 authentication. The same applies to routers adjacent to interfaces using MD5 authentication. Do not stop and start OMPROUTE without waiting for this required time interval.

Rereading the configuration file

The MODIFY *procname*,RECONFIG command is used to reread the OMPROUTE configuration file. This command accepts and applies only statements of the following types that have been added to the configuration file since the last time the file was read:

- OSPF_INTERFACE
- RIP_INTERFACE
- INTERFACE
- IPV6_OSPF_INTERFACE
- IPV6_OSPF (ROUTERID parameter only)
- IPV6_RIP_INTERFACE
- IPV6_INTERFACE

These configuration statements must be reread from the configuration file using the MODIFY *procname*,RECONFIG command before the interfaces to which they refer are defined to the TCP/IP stack. If you have coded GLOBAL_OPTIONS IGNORE_UNDEFINED_INTERFACES=YES in your OMPROUTE configuration file, you can use OMPROUTE reconfiguration to add a definition for an interface that has already been configured in the stack but ignored by OMPROUTE. However, OMPROUTE does not associate the interface with the new definition until the interface has been deleted from the stack and re-added.

Restriction: The IPV6_OSPF statement (ROUTERID parameter only) is recognized by this command only if there is no existing IPv6 OSPF router ID value set in OMPROUTE. This command cannot be used to change the value of an existing IPv6 OSPF router ID.

Enabling or disabling the OMPROUTE subagent

Use the MODIFY <procname>,ROUTESA=ENABLE command or the MODIFY <procname>,ROUTESA=DISABLE command to enable or disable the OMPROUTE subagent.

Rule: To change any other value on the ROUTESA_CONFIG statement, the OMPROUTE application must be recycled.

The OMPROUTE subagent implements RFC 1850, *OSPF Version 2 Management Information Base*, for the OSPF protocol. The ROUTESA_CONFIG statement is used in the OMPROUTE configuration file to configure the OMPROUTE subagent. For details on ROUTESA_CONFIG, see z/OS Communications Server: IP Configuration Reference.

Changing the cost of OSPF links

The cost of an OSPF interface can be dynamically changed using the MODIFY <procname>,OSPF,WEIGHT,NAME=<if_name>,COST=<cost> command. The cost of an IPv6 OSPF interface can be dynamically changed using the MODIFY <procname>,IPV6OSPF,WEIGHT,NAME=<if_name>,COST=<cost> command. This new cost is flooded quickly throughout the OSPF routing domain, and modifies the routing immediately.

The cost of the interface reverts to its configured value whenever OMPROUTE is restarted. To make the cost change permanent, you must change the appropriate OSPF_INTERFACE or IPV6_OSPF_INTERFACE statement in the configuration file.

Controlling OMPROUTE tracing and debugging

The following commands are used to start, stop, or change the level of OMPROUTE tracing and debugging:

- MODIFY <procname>,TRACE=n : for OMPROUTE tracing for initialization and IPv4 routing protocols; n can be 0–2.
- MODIFY <procname>,DEBUG=n : for OMPROUTE debugging for initialization and IPv4 routing protocols; n can be 0–4.
- MODIFY <procname>,SADEBUG=n : for OMPROUTE subagent debugging; n can be 0 or 1.
- MODIFY <procname>,TRACE6=n : for OMPROUTE tracing for IPv6 routing protocols; n can be 0–2.
- MODIFY <procname>,DEBUG6=n : for OMPROUTE debugging for IPv6 routing protocols; n can be 0–4.

Tip: Use of OMPROUTE tracing and debugging affects OMPROUTE performance and might require increasing the Dead_Router_Interval on OSPF interfaces to keep neighbor adjacencies from collapsing.

Steps for configuring OSPF and RIP (IPv4 and IPv6)

RIP uses a distance vector algorithm to calculate the best path to a destination based on the number of hops in the path. OSPF uses a link state or shortest path first algorithm.

Before you begin

The following steps are the general steps to configure OSPF and RIP. Details follow the general steps.

1. Set the global configuration options.
2. Set the OSPF router ID, if the OSPF protocol is used.
3. Define OSPF areas, if the OSPF protocol is used.
4. Limit information exchange between OSPF areas, if the OSPF protocol is used.
5. Define IPv4 interfaces, if the IPv4 OSPF or IPv4 RIP protocol is used.
6. Define IPv6 interfaces, if the IPv6 OSPF or IPv6 RIP protocol is used.
7. Define interface costs (OSPF_INTERFACE, RIP_INTERFACE, IPV6_OSPF_INTERFACE, and IPV6_RIP_INTERFACE).
8. Configure virtual links, if the OSPF protocol is used.
9. Manage high-cost links, if the OSPF protocol is used.
10. Define RIP filters, if the RIP protocol is used.
11. Define route precedence in a multiprotocol environment, if the OSPF protocol is used.

If policy-based routing is used on the TCP/IP stack and the routing policy is configured with dynamic routing parameters, additional configuration to OMPROUTE is not needed for dynamic routing support to be provided for the policy-based route tables. The dynamic routing parameters that are specified in the routing policy are provided to OMPROUTE by the TCP/IP stack to control the scope of dynamic routes that are computed by OMPROUTE. For a description of this function, see "Policy-based routing" on page 349.

Procedure

Perform the following steps to configure OSPF and RIP (IPv4 and IPv6):

1. Set the global configuration options.

   Use the GLOBAL_OPTIONS statement to set the Ignore_Undefined_Interfaces parameter for OMPROUTE processing. Undefined interfaces are stack interfaces that are not configured by using OMPROUTE interface definitions, either explicit definitions or wildcard definitions. If you specify the value YES on the Ignore_Undefined_Interfaces parameter (Ignore_Undefined_Interfaces=YES), OMPROUTE does not configure undefined interfaces with default values. When Ignore_Undefined_Interfaces is set to the value NO, OMPROUTE configures undefined interfaces with default values that can override the stack definition values. The default value for Ignore_Undefined_Interfaces is NO.

   Tip: Specify Ignore_Undefined_Interfaces=YES to ensure that undefined interfaces are not configured with default values that might affect network connectivity.

2. Set the OSPF router ID, if the OSPF protocol is used.

   Every router in an OSPF autonomous system must be assigned a unique router ID.

   - IPv4 OSPF

     Code the ROUTERID parameter of the OSPF configuration statement within the OMPROUTE configuration file to assign the router ID. The value must be the IP address of one of the OSPF_INTERFACEs defined in the OMPROUTE configuration file. If the ROUTERID parameter of the OSPF configuration statement is not coded, OMPROUTE chooses the IP address from one of the OSPF_INTERFACE statements as the router ID. Because dynamic VIPAs (DVIPAs) can move between z/OS hosts within a sysplex, the router ID should be the IP address of a physical interface or a static VIPA, but not a dynamic VIPA. For more information about the rules and guidelines for the ROUTERID parameter of the OSPF statement, see z/OS Communications Server: IP Configuration Reference.

     In the example network that is shown in Figure 48 on page 282, the router ID is set to the static VIPA address that represents each OMPROUTE router. TCPCS4 has ROUTERID=4.4.4.4, and TCPCS7 has ROUTERID=7.7.7.7.

   - IPv6 OSPF

     The ROUTERID parameter of the IPV6_OSPF configuration statement is coded within the OMPROUTE configuration file to assign the IPv6 OSPF router ID. This router ID is any IPv4-style dotted decimal value except for 0.0.0.0, with care taken to assure its uniqueness across routers within the IPv6 autonomous system. If the ROUTERID parameter of the IPV6_OSPF configuration statement is not coded and the IPv4 OSPF protocol is also being used, OMPROUTE uses the IPv4 OSPF router ID as the IPv6 OSPF router ID. If the IPv4 OSPF protocol is not being used, the ROUTERID parameter of the IPV6_OSPF statement must be specified. For more information about the rules and guidelines for the ROUTERID parameter of the IPV6_OSPF statement, see z/OS Communications Server: IP Configuration Reference.

     In the example network that is shown in Figure 49 on page 283, the ROUTERID parameter of the IPV6_OSPF statement on TCPCS64 is to be explicitly configured by using ROUTERID=64.64.64.64.

3. Define OSPF areas, if the OSPF protocol is used.

   The sample network that is shown in Figure 48 on page 282 and Figure 49 on page 283 depicts an IPv4 network and an IPv6 network, both divided by using two methods. The first division is between IP subnetworks within the OSPF autonomous system (AS) and IP subnetworks external to the OSPF AS (within the RIP AS). The subnetworks included within each OSPF AS are further subdivided into regions called areas. OSPF areas are collections of contiguous IP subnetworks. The function of areas

is to reduce the OSPF overhead that is required to compute routes to destinations in different areas. Overhead is reduced because less information is exchanged and stored by routers and because fewer processor cycles are required for a less complex route table calculation.

Every OSPF AS with multiple areas must have at least a backbone area. The backbone is always identified by area number 0.0.0.0. For networks with multiple areas, the backbone provides a core that connects the areas. Unlike other areas, the backbone subnets can be physically separate. In this case, logical connectivity of the backbone is maintained by configuring virtual links between backbone routers across intervening non-backbone areas. For more information, see step "8" on page 320.

Routers that attach to more than one area function as area border routers. All area border routers are part of the backbone, so they must either attach directly to a backbone IP subnet or be connected to another backbone router over a virtual link.

The information and algorithms that are used by OSPF to calculate routes vary according to whether the destination is within the same area, in a different area within the OSPF AS, or external to the OSPF AS. Every router maintains a database of all links within its area. A shortest path first algorithm is used to calculate the best routes to destinations within the area from this database. Routes between areas are calculated from summary advertisements that are originated by area border routers for destinations that are located in other areas of the OSPF AS. External routes (for example, routes to destinations that lie within a RIP AS) are calculated from AS external advertisements that are originated by AS boundary routers and flooded throughout the OSPF AS.

- IPv4 OSPF

  Use the AREA configuration statement to define the areas to which a router attaches. If you do not use the AREA statement, the default is that all OSPF interfaces attach to the backbone area. In the sample network, TCPCS4 and TCPCS7 are both area border routers that belong to both the backbone area (0.0.0.0) and area 1.1.1.1.

```
AREA
  Area_Number=0.0.0.0;

AREA
  Area_Number=1.1.1.1;
```

- IPv6 OSPF

  Use the IPV6_AREA configuration statement to define the areas to which a router attaches. If you do not use the IPV6_AREA statement, the default is that all IPv6 OSPF interfaces attach to the backbone area. In the sample network, TCPCS64 and TCPCS67 are area border routers that belong to both the backbone area (0.0.0.0) and area 6.6.6.6.

```
IPV6_AREA
   Area_Number=0.0.0.0;
IPV6_AREA
   Area_Number=6.6.6.6;
```

4. Limit information exchange between OSPF areas, if the OSPF protocol is used.

   When area border routers are configured, the OSPF route information that crosses the area boundary can be controlled with configuration statements. For information about the usefulness of multiple areas in the z/OS CS environment, see "Minimizing the routing responsibility of z/OS Communications Server" on page 324.

   One option is to define an area as a stub area. AS external advertisements are never flooded into stub areas. In addition, the origination into the stub area of summary advertisements for interarea routes can be suppressed, creating what is commonly known as a *totally stubby* area.

   Destinations external to the stub area are still reachable since the area border routers advertise default routes into stub areas. Traffic within the stub area for unknown destinations is forwarded to the area border router (by using the default route). It is acceptable for routers within the area to use a default route for traffic that is destined outside the AS. The border router uses its more complete routing information to forward the traffic on an appropriate path toward its destination.

The following requirements must be met for an area to be defined as a stub area:

- No virtual links are configured through the area to maintain backbone connectivity.
- No routers within the area are AS boundary routers (OSPF routers that advertise routes from external sources as AS external advertisements).

  Tip: Static routes and RIP routes are AS external.

Another option is to use IP address ranges to limit the number of summary advertisements that are originated out of an area. In IPv4, a range is defined by an IP address and an address mask. Destinations are considered to fall within the range when the destination address and the range IP address match after the range mask is applied to both addresses. In IPv6, a range is defined by an IP address prefix and a prefix length, and destinations are considered to fall within the range if a destination address and a range IP address match for the length of the range prefix.

When a range is configured for an area at an area border router, the border router suppresses summary advertisements for destinations within that area that fall within the range. The suppressed advertisements would have been originated into the other areas to which the border router attaches. Instead, the area border router originates a single summary advertisement for the range or no advertisement at all, depending on the option that is chosen with the configuration statement.

Rules:

a. If the range is not advertised, there will be no interarea routes for any destination that falls within the range.

b. Ranges cannot be used for areas through which virtual links are configured to maintain backbone connectivity.

- IPv4 OSPF

  The following statement configures an OSPF area as a stub area. The Import_Summaries=No parameter results in the suppression of summary advertisements for interarea routes into the stub area, creating a totally stubby area:

```
AREA
  Area_Number=2.2.2.2
  Stub_area=Yes
  Import_Summaries=No;
```

  In the sample network that is shown in Figure 48 on page 282, the following RANGE statement could be configured on TCPCS7 to prevent TCPCS7 from advertising destinations in the 9.67.101.0 subnet into the backbone area (Area 0.0.0.0):

```
RANGE
   IP_Address=9.67.101.0
   Subnet_Mask=255.255.255.0
   Area_Number=1.1.1.1
   Advertise=No;
```

- IPv6 OSPF

  The following statement configures an IPv6 OSPF area as a stub area. The Import_Summaries=No parameter results in the suppression of summary advertisements for interarea routes into the stub area, creating a totally stubby area:

```
IPV6_AREA
   Area_Number=1.1.1.1
   Stub_area=Yes
   Import_Prefixes=No;
```

In the sample network, the following IPV6_RANGE statement could be configured on TCPCS67 to cause TCPCS67 to advertise all destinations in the 2001:0DB8:0:31::/64 prefix into the backbone area (Area 0.0.0.0) as a single route to the 2001:0DB8:0:31::/64 prefix:

```
IPV6_RANGE
   Prefix=2001:0DB8:0:31::/64
   Area_Number=6.6.6.6
   Advertise=Yes;
```

5. Define IPv4 interfaces, if the IPv4 OSPF or IPv4 RIP protocol is used.

   Define each interface in use by the stack to OMPROUTE by using an OSPF_INTERFACE, RIP_INTERFACE, or INTERFACE statement. This topic describes the differences between interface types for your consideration when you are configuring interfaces to OMPROUTE. In general, use the following guidelines:

   - An interface over which the OSPF protocol is communicated with other routers must be configured with the OSPF_INTERFACE statement.
   - An interface over which the RIP protocol is communicated with other routers must be configured with the RIP_INTERFACE statement.
   - Either configure all other interfaces with the INTERFACE statement, or configure OMPROUTE to ignore undefined interfaces by using the IGNORE_UNDEFINED_INTERFACES parameter of the GLOBAL_OPTIONS statement in the OMPROUTE configuration file. It is important that one or the other is done.

     If OMPROUTE is not configured to ignore undefined interfaces, it configures stack interfaces that are not defined to OMPROUTE with default values. These values might not be desirable. For example, the class mask is used as the subnet mask and 576 is used as the MTU value. Furthermore, OMPROUTE overrides stack values with its default values. To prevent this situation from happening, either configure every interface, even those interfaces that are not using RIP or OSPF, or configure OMPROUTE to ignore undefined interfaces.

   A VIPA interface is an exception to these guidelines and is described in more detail in this step.

   z/OS Communications Server enforces RFC rules against using either a subnetwork broadcast or network address as a host address. (An address that has all ones in the host portion is a subnet broadcast address. An address that has all zeros in the host portion is the subnet network address.) Therefore, the subnet_mask on an OSPF_INTERFACE, RIP_INTERFACE, or INTERFACE statement should have enough zero bits such that no home address in that subnet has all zeros or all ones in the host portion of the address. For example, if a subnet has two home addresses 10.1.1.1 and 10.1.1.2, the subnet mask must have zeros in at least 2 bits; for example, 255.255.255.252. However, if a subnet has four home addresses 10.1.1.1, 10.1.1.2, 10.1.1.3, and 10.1.1.4, the subnet mask must have zeros in at least 3 bits; for example, 255.255.255.248. In this case, there can be up to six home addresses in that subnet (10.1.1.1 through 10.1.1.6). In general, if a subnet mask has n zero bits, then there can be up to ((2**n)-2) home addresses in that subnet. This limit applies even if the home addresses are configured on different TCP/IP stacks.

   Rules:

   a. OMPROUTE supports a maximum of 255 real, physical, IPv4 interfaces (that is, interfaces on which data can actually be sent and received). There is no theoretical limit on how many VIPAs can be configured, though there are practical limits that are imposed by network design.
   b. RIP version 1 uses broadcast and RIP version 2 uses multicast. RIP version 1 cannot run on a medium that supports multicast but not broadcast, which is the default QDIO configuration. To configure the OSA-Express feature operating in QDIO mode (for example, gigabit Ethernet) to send and receive broadcast packets, use the IPBCAST parameter on the LINK statement in your TCP/IP profile. The OSA-Express microcode level must support broadcast to use this parameter. Some levels of OSA-Express microcode support multicasting but not broadcasting. In this case, RIP version 2, which relies on multicast, is preferred over RIP version 1, which relies on broadcast.

   - Configuring multi-access parallel interfaces

Whenever configuring multi-access parallel interfaces (primary and backup redundant interfaces with IP addresses in the same network) for OMPROUTE (OSPF), use the Parallel_OSPF parameter on the OSPF_INTERFACE statement to designate whether each OSPF interface is primary or backup. If the IP_address parameter on an OSPF_INTERFACE statement is using wildcards (*), also include the Name parameter for the interface to distinguish the primary from the backups. For more information about the OSPF_INTERFACE statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

- Point-to-point (For example, CTC)

  For point-to-point interfaces, the destination IP address must be known to OMPROUTE. If OSPF or RIP is run on the interface, the destination IP address is learned when the router at the other end comes up and shares information with OMPROUTE. Additionally, defining a destination address on an interface that is running OSPF or RIP allows OMPROUTE to learn and advertise a route to that destination address before a neighboring router becomes fully adjacent. This can be beneficial if the neighboring router takes a long time to come up, or is otherwise not expected to be promptly and reliably available.

  Tip: For a point-to-point interface that is running OSPF, OMPROUTE does not advertise a host route to its own home address on the point-to-point interface. By default, OMPROUTE advertises a host route to the link destination, and relies on the router at the other end to advertise a host route to the OMPROUTE home address. This behavior is described by the OSPF architecture, RFC 1583 (OSPF version 2), section 12.4.1. This means that the OMPROUTE home address on the point-to-point link cannot be advertised as reachable unless there is a neighboring router available to advertise it. OMPROUTE implements an extension that is described in RFC 2328 to overcome this limitation. If a neighboring router will not be reliably available over the point-to-point link, you might want to code the parameter SUBNET=YES on the OSPF_INTERFACE statement for the point-to-point interface. This parameter causes OMPROUTE to implement option 2 described in RFC 2328, section 12.4.1.1, and advertise a route to the point-to-point link subnet address, which makes both endpoints reachable regardless of the status of a neighboring router.

  If the interface is only an INTERFACE (not running OSPF or RIP), specify the DESTINATION_ADDR parameter to allow for the creation of a host route to the address at the remote end of the interface.

  Sample OSPF_INTERFACE:

```
OSPF_INTERFACE
    IP_Address=9.67.106.7
    Name=CTC7TO4
    Subnet_mask=255.255.255.0
    Attaches_to_Area=1.1.1.1
    Destination_Addr=9.67.106.4;
```

  Sample RIP_INTERFACE:

```
RIP_INTERFACE
    IP_Address=9.67.103.7
    Name= CTC7TO6
    Subnet_mask=255.255.255.0
    Destination_Addr=9.67.103.6
    RIPV2=Yes;
```

Sample INTERFACE:

```
INTERFACE
    IP_Address=9.67.111.1
    Name=CTCX
    Subnet_mask=255.255.255.0
    Destination_addr=9.67.111.2;
```

- Point-to-multipoint

  For point-to-multipoint capable interfaces (for example MPCPTP interfaces including XCF and IUTSAMEH connections), OMPROUTE must know the IP addresses of the other routers (neighbors) with which it must communicate the OSPF or RIP packets. However, because of underlying signaling that takes place when a host connects to these network types, the stack is able to learn the required addresses. In turn, OMPROUTE learns those IP addresses from the stack. As a result, it is not necessary to configure the IP addresses of the other routers on the interface statements.

  Sample OSPF_INTERFACE:

  ```
  OSPF_INTERFACE
      IP_Address=9.27.13.81
      Name=XCFD00
     Attaches_to_Area=1.1.1.1
      Subnet_mask=255.255.255.0;
  ```

  Sample RIP_INTERFACE:

  ```
  RIP_INTERFACE
      IP_Address=9.27.23.81
      Name=MPCA01
      Subnet_mask=255.255.255.0
      RIPV2=Yes;
  ```

  Sample INTERFACE:

  ```
  INTERFACE
      IP_Address=9.27.33.81
      Name=XCFB00
      Subnet_mask=255.255.255.0;
  ```

- Broadcast network interfaces (For example, Ethernet)

  When the OSPF or RIP protocol is communicated over a broadcast medium such as Ethernet, these networks allow for broadcasting and multicasting. Therefore, it is not necessary for OMPROUTE to know the IP addresses of the other routers on the network for OSPF or RIP packets to be communicated with those routers. OMPROUTE sends packets to the other routers on the network by using appropriate broadcast or multicast addresses. The IP addresses of the other routers are learned as OSPF/RIP packets are received from them. The OSPF_INTERFACE must include the ROUTER_PRIORITY parameter to help elect a designated router for the network.

  Sample OSPF_INTERFACE:

  ```
  OSPF_INTERFACE
      IP_Address=9.59.101.5
      Name=ETH1
     Subnet_mask=255.255.255.0
      Attaches_to_Area=1.1.1.1
      Cost0=2
      Router_Priority=1;
  ```

Sample RIP_INTERFACE:

```
RIP_INTERFACE
    IP_Address=9.29.107.3
    Name=ETH2
    Subnet_mask=255.255.255.0
    RIPV2=Yes;
```

Sample INTERFACE:

```
INTERFACE
    IP_Address=9.77.14.49
    Name=ETHB00
    Subnet_mask=255.255.255.0;
```

For interfaces into broadcast media which contain routers that do not support multicast, it is possible to configure the interfaces as non-broadcast network interfaces, which would cause OMPROUTE to unicast to the neighbor addresses rather than using a multicast address. However, it would also be necessary to configure all the routers on the network to unicast. Otherwise, their multicast packets would never be received.

Note that it is possible to define neighbors by using DR_NEIGHBOR and/or NO_DR_NEIGHBOR parameters for OSPF_INTERFACEs and by using NEIGHBOR parameters for RIP_INTERFACEs that are broadcast capable, but it is not required or advised. If you define neighbors on these interfaces, you must define all of them, as OMPROUTE does not communicate RIP or OSPF to undefined neighbors if any are defined on an interface.

- VIPA interfaces (Static VIPA and Dynamic VIPA)

  If only the RIP protocol is used by OMPROUTE, define VIPA interfaces with the INTERFACE statement. If only OSPF or if both OSPF and RIP are used by OMPROUTE, define VIPA interfaces with the OSPF_INTERFACE statement.

  Sample OSPF_INTERFACE:

  OSPF example:

  ```
      OSPF_INTERFACE
      IP_Address=4.4.4.4
      Name=VIPA1
     Subnet_mask=255.255.255.252;
  ```

  Sample INTERFACE:

  non-OSPF example:

  ```
      INTERFACE
      IP_Address=6.6.6.6
      Name=VIPA1
      Subnet_mask=255.255.255.252;
  ```

  Rule: The most specific subnet mask you can specify is 255.255.255.252.

  If the name in an OSPF_INTERFACE or INTERFACE statement refers to a link of type VIRTUAL, then OMPROUTE generates and advertises the following routes whenever applicable:

  – A network route to the network specified in that statement
  – A subnet route to the subnet specified in that statement
  – A host route to the IP_address specified in that statement

  The following conditions apply for advertising these routes on a physical network interface to a network:

  – Network route - if VIPA is not in the same network as the physical network interface and is allowed by filters or RANGE.

– Subnet route - VIPA subnet routes are advertised in OMPROUTE in all conditions, except for RIP when filters prevent it.
– Host route - as allowed by filters or RANGE. Advertisement of the host route for a VIPA defined on an OSPF_INTERFACE statement can be controlled by the SUBNET parameter on the OSPF_INTERFACE statement that defines that VIPA. If SUBNET=YES, then the host route is not advertised. If SUBNET=NO (the default), the host route is advertised. Take care in using this parameter. Do not suppress VIPA host routes for dynamic VIPAs or for VIPAs whose subnet might exist on multiple hosts. It is up to the user to ensure that these restrictions are enforced, as they are not and cannot be enforced by OMPROUTE.

On the RIP_INTERFACE statement for a physical network interface, the VIPA routes are allowed to be advertised by the following filter parameters:

– Send_Net_Routes
– Send_Subnet_Routes
– Send_Host_Routes, and Send_Only

In addition, the global FILTER and Send_Only statements for RIP can be used to specify which routes are advertised or not.

For OSPF, the RANGE statement can be used to advertise or not to advertise the VIPA routes external to an area in terms of address range based on a subnet mask.

Note: For RIP, the Send_Only = (VIRTUAL) filter with the Send_Net_Routes, Send_Subnet_Routes, and Send_Host_Routes filters, or the FILTER statement with VIPA routes, indicates whether VIPA routes can be advertised over a RIP interface. Unlike RIP, there are no routing filters for OSPF. For OSPF, the RANGE statement can be used to control which address range of routes can be advertised or not advertised external to an area; however, it is not granular enough for use as a routing filter. In area-border router configurations, if there are multiple VIPA addresses that are uniquely subnetted, the RANGE statement can be used to specify which VIPA subnet address range of routes can be advertised or not advertised external to an area.

For Dynamic VIPA (DVIPA), link names are assigned programmatically by the stack when the DVIPA is created. Therefore, the name field that is set on the INTERFACE or OSPF_INTERFACE statement is ignored by OMPROUTE for DVIPAs.

Because a stack can have many DVIPAs defined, and DVIPA ranges, more wildcard capabilities exist on the OSPF_INTERFACE and INTERFACE statements for use only with DVIPAs.

Ranges of DVIPA interfaces can be defined by using the Subnet_Mask parameter on the OSPF_INTERFACE or INTERFACE statement. The range that is defined in this way includes all the IP addresses that fall within the subnet that is defined by the mask and the IP address. The IP address parameter must be the subnet number of the range that is being defined, not a host address within that range. Further information about the Subnet_Mask parameter is available earlier in this step.

In the following example, DVIPA interfaces in the range of 10.138.65.81 through 10.138.65.94 are defined:

Sample OSPF_INTERFACE:

OSPF example:

```
   OSPF_INTERFACE
   IP_Address=10.138.65.80
   Name=DVIPAs
  Subnet_mask=255.255.255.240;
```

Sample INTERFACE:

non-OSPF example:

```
   INTERFACE
   IP_Address=10.138.65.80
```

```
    Name=DVIPAs
    Subnet_mask=255.255.255.240;
```

In the following example, only an interface with home address 10.138.65.98 is being defined, because the subnet number (obtained by using a binary file AND operation of the IP_Address parameter and the Subnet_mask parameter) for this definition is 10.138.65.96. Because the IP_Address parameter does not equal that subnet number, this definition cannot be treated as a DVIPA wildcard.

```
OSPF_INTERFACE
IP_Address=10.138.65.98
Name=DVIPA
Subnet_mask = 255.255.255.240;
```

In the following example, DVIPAs with IP addresses 10.138.120.x and 10.138.121.x with a 24-bit mask (255.255.255.0) result in two subnets 10.138.120.0/24 and 10.138.121.0/24. The 10.138.120.0/24 subnet ranges from 10.138.120.1 through 10.138.120.254 and the 10.138.121.0/24 subnet ranges from 10.138.121.1 through 10.138.121.254. Two OSPF_INTERFACE statements with the extended wildcard capabilities are defined.

```
OSPF_INTERFACE
   IP_Address=10.138.120.0
   Name=DVIPAs
   Subnet_mask=255.255.255.0;
   IP_Address=10.138.121.0
   Name=DVIPAs
   Subnet_mask=255.255.255.0;
```

In the following example, DVIPAs with IP addresses 10.138.120.x and 10.138.121.x with a 23-bit mask (255.255.254.0) result in a collapsed subnet 10.138.120.0/23. The 10.138.120.0/23 subnet ranges from 10.138.120.1 through 10.138.121.254.

```
OSPF_INTERFACE
   IP_Address=10.138.120.0
   Name=DVIPAs
   Subnet_mask=255.255.254.0;
```

Notes:

– If the conventional wildcard IP_Address=10.138.120.* is coded, OMPROUTE does not match the 10.138.121.x addresses for the DVIPAs because the subnet mask is not taken into account. Therefore, these DVIPAs might not be added to the OSPF domain.

– If DVIPAs across multiple networks are collapsed into one supernet, the same principle with the mask applies to the extended wildcard capability. For example, a supernet address 192.168.100.0 with a 23-bit mask (255.255.254.0) covers the range 192.168.100.1 through 192.168.101.254.

You must consider an additional issue when VIPAs are being moved between TCP/IP stacks and dynamic routing is provided for those stacks by OMPROUTE. This movement of VIPAs can be done manually or automatically with Dynamic VIPAs. For the VIPAs to be correctly processed and advertised by the routing protocols, they (like all other interfaces) must be configured to OMPROUTE at the time that they become active on the TCP/IP stack. This configuration of VIPAs to OMPROUTE can be accomplished by:

– Explicitly configuring each VIPA with its own OSPF_INTERFACE or INTERFACE statement

– Configuring a range of DVIPAs with a single OSPF_INTERFACE or INTERFACE statement

– Configuring a group of VIPAs with a single OSPF_INTERFACE or INTERFACE statement, by using the wildcard feature available on the interface statements

The advised approach for configuring OMPROUTE for VIPAs that might move is to preconfigure the OMPROUTE on each TCP/IP stack with all VIPAs that can potentially exist on that stack at some time. Preconfiguring in this way prepares each OMPROUTE for the possible addition of the

VIPAs to its stack. During times when the VIPAs do not exist on a particular OMPROUTE stack, the configuration information will not be used. However, during periods when the VIPAs do exist on that OMPROUTE stack, the configuration information is available for use by OMPROUTE. This method is advised because of its ability to respond to movement of the VIPAs between TCP/IP stacks without modification of the OMPROUTE configuration with each move.

If the pre-configuration of VIPAs described in this topic is not done, it is still possible to define a VIPA to OMPROUTE such that it is properly processed and advertised when it becomes active on the corresponding TCP/IP stack. To do this configuration, add the appropriate OSPF_INTERFACE or INTERFACE statement to the OMPROUTE configuration file and then cause OMPROUTE to reread the configuration file by issuing the MODIFY *procname*,RECONFIG command.

Rule: If you did not code GLOBAL_OPTIONS IGNORE_UNDEFINED_INTERFACES=YES in your OMPROUTE configuration file, you must modify the OMPROUTE configuration file and issue the RECONFIG command before the first time that the VIPA moves to the corresponding TCP/IP stack.

Guideline:

If you did code GLOBAL_OPTIONS IGNORE_UNDEFINED_INTERFACES=YES in your OMPROUTE configuration file, you can use OMPROUTE reconfiguration to add a definition for a VIPA interface that is or was active in the stack but ignored by OMPROUTE.

However, OMPROUTE does not associate the VIPA with the new definition until the interface is deleted from the stack and readded, either by DVIPA movement or by some other method.

Method of assigning interface definitions to stack interfaces (wildcard and explicit):

Wildcard interface definitions can be a convenient way of making your interface definitions easier. However, to avoid unintended results, be sure to understand how they are parsed, and how different types of interface definitions interact with each other. The following outline shows the algorithm that OMPROUTE uses to find the matching definitions in the OMPROUTE configuration file for an IPv4 stack interface.

a. Search for a RIP_Interface definition for the interface as follows:

   i) Search for an explicit matching RIP_Interface definition (IP address matches exactly if a dynamic VIPA, otherwise name and IP address match exactly). If found, use that definition and go to step b.

   ii) If the interface is a dynamic VIPA, search for a RIP_Interface definition that matches as a dynamic VIPA wildcard. A matching dynamic VIPA wildcard definition is one where the definition IP address matches the value that is obtained by ANDing the definition subnet mask with the interface home address. The definition name parameter is ignored during a search for a dynamic VIPA wildcard definition. If a matching dynamic VIPA wildcard definition is found, use that definition and go to step b.

   iii) Search for a one-octet wildcard RIP_Interface definition (n.n.n.*), where the first three octets match the first three octets of the interface home address and the name matches the interface link name. If found, use that definition and go to step b.

   iv) Search for a two-octet wildcard RIP_Interface definition (n.n.*.*), where the first two octets match the first two octets of the interface home address and the name matches the interface link name. If found, use that definition and go to step b.

   v) Search for a three-octet wildcard RIP_Interface definition (n.*.*.*), where the first octet matches the first octet
   of the interface home address and the name matches the interface link name. If found, use that definition and go to step b.

   vi) Search for a one-octet wildcard RIP_Interface definition (n.n.n.*), where the first three octets match the first three octets of the interface home address, ignoring the name parameter when coded. If found, use that definition and go to step b.

   vii) Search for a two-octet wildcard RIP_Interface definition (n.n.*.*), where the first two octets match the first two octets of the interface home address, ignoring the name parameter when coded. If found, use that definition and go to step b.

viii) Search for a three-octet wildcard RIP_Interface definition (n.*.*.*), where the first octet matches the first octet
of the interface home address, ignoring the name parameter when coded. If found, use that definition and go to step b.

ix) If there is a four-octet wildcard RIP_Interface definition (*.*.*.* or ALL), use that definition and go to step b.

Restriction: Only 1 four-octet wildcard of each type (OSPF_INTERFACE or RIP_INTERFACE) is allowed.

b. Search for an OSPF_Interface definition for the interface. This step is done regardless of the outcome of step a. The steps for searching OSPF_Interface definitions are the same as the steps for searching RIP_Interface definitions, except that OSPF_Interface definitions are searched.

c. If either a RIP_Interface or an OSPF_Interface definition, or both, are found, the algorithm is complete. In this case, Interface definitions are not searched. If neither a RIP_Interface or an OSPF_Interface definition was found, go to step d.

d. Search for an Interface definition for the interface. The steps for searching Interface definitions are the same as the steps for searching RIP_Interface statements, except that Interface definitions are searched.

e. If no definitions are found, check the value of Global_Options Ignore_Undefined_Interfaces. If this option is turned on, the interface is ignored. If it is not turned on, the interface is treated as if it were defined with an Interface statement, with an MTU value of 576 and a subnet mask of the class mask.

The algorithm is complete. Key conclusions of this algorithm are as follows:

- A wildcard interface definition with a matching name is preferred over a wildcard interface definition of the same type without a matching name, regardless of which definition is a more specific wildcard.
- If the interface is not a dynamic VIPA, the subnet mask that is coded on the definition statement has no influence on which definition, wildcard or otherwise, is selected for an interface. The subnet mask is a characteristic of the definition, not a selection criterion.
- If a RIP_Interface or an OSPF_Interface definition, or both, are found, Interface definitions are not considered. This means that any matching RIP_Interface or OSPF_Interface definition supersedes all Interface definitions, even if the Interface definitions are explicit or are more specific wildcard matches. For example, a three-octet wildcard OSPF_Interface definition supersedes an explicit Interface definition.
- An interface can be both a RIP_Interface and an OSPF_Interface. OMPROUTE supports running both protocols over the same interface. However, an interface cannot be both an interface that runs no routing protocol (that is, defined with the Interface statement) and one that runs RIP, OSPF, or both.

6. Define IPv6 interfaces, if the IPv6 OSPF or IPv6 RIP protocol is used.

   Each IPv6 interface in use by the stack can be defined to OMPROUTE by using an IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, or IPV6_INTERFACE statement. Defining IPv6 interfaces to OMPROUTE is much simpler than defining IPv4 interfaces because you do not specify IP addresses or MTU sizes to OMPROUTE. Instead, you define interfaces by their names and OMPROUTE learns IP addresses and MTU sizes from the TCP/IP stack. Also, because multicast support is a basic requirement of IPv6, there are no non-broadcast multiaccess considerations or other considerations that might require definitions of neighbors or destination addresses.

   Use the following guidelines when you are configuring IPv6 interfaces to OMPROUTE:

   - An interface over which the IPv6 OSPF protocol is communicated with other routers must be configured with the IPV6_OSPF_INTERFACE statement.
   - An interface over which the IPv6 RIP protocol is communicated with other routers must be configured with the IPV6_RIP_INTERFACE statement.

- All other interfaces can be configured with the IPV6_INTERFACE statement when OMPROUTE default values are not acceptable or you want to define more prefixes on the interface.
- The interface name must be coded on the IPV6_INTERFACE, IPV6_OSPF_INTERFACE, and IPV6_RIP_INTERFACE statements. The value of the NAME parameter must match the interface name that is coded on the INTERFACE statement in the TCP/IP profile. Wildcard names that end with an asterisk (*) can be coded. For example, OSAQDIO* would match stack interfaces named OSAQDIO1, OSAQDIO2, OSAQDIOABC, and so on.

  If the interface is dynamically generated by the TCP/IP stack, its name parameter must match what is generated by the TCP/IP stack. Interfaces that are dynamically generated by the TCP/IP stack are named as follows:

  – An IPv6 dynamic XCF interface that is generated to attach to other TCP/IP stacks within the same z/OS host is always named EZ6SAMEMVS.
  – An IPv6 dynamic XCF interface that is generated to attach to TCP/IP stacks in another z/OS image is always named EZ6XCF*xx*, where *xx* is the SYSCLONE value of the other z/OS host.

  If the routing parameters for your dynamic XCF interfaces are all to be the same, you can use wildcard definitions to avoid having to know and build definitions for every possible dynamic XCF interface on your system. For example, a wildcard definition for name EZ6* would match all dynamic XCF interfaces that can be generated on a TCP/IP stack. A wildcard definition for EZ6XCF* would match all dynamic XCF interfaces that can be generated to attach to other z/OS images.

  The following definition would define all IPv6 dynamic XCF interfaces to OMPROUTE, with the hello and dead router intervals changed from the default values:

```
IPV6_OSPF_INTERFACE
    NAME=EZ6*
    HELLO_INTERVAL = 30
    DEAD_ROUTER_INTERVAL = 120;
```

  If the default values of 10 and 40 for the hello and dead router intervals are acceptable, this definition can be simplified even more:

```
IPV6_OSPF_INTERFACE
    NAME=EZ6*;
```

- To define one or more prefixes on an interface, use the PREFIX parameter on the IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, and IPV6_INTERFACE statements. You must use the PREFIX parameter only for prefixes that you must define to an interface, which cannot be learned using IPv6 router discovery. Also note that prefixes defined to OMPROUTE in this manner are not used by TCP/IP to autoconfigure home addresses on the interface.

  The following sample shows an IPv6 OSPF interface with prefixes defined:

```
IPV6_OSPF_INTERFACE
    NAME=OSAQDIO4L6
    PREFIX=2001:0DB8:1::/48
    PREFIX=2001:0DB8:2::/48;
```

  The prefixes defined in this manner on an IPv6 OSPF interface are advertised as reachable, and are also included in the link LSA generated by OMPROUTE, so all IPv6 OSPF routers on the link will know they are local prefixes. If OMPROUTE is also running IPv6 RIP, they are also advertised into the IPv6 RIP autonomous system as reachable, if IPv6 RIP filters permit it.

  The following sample shows an IPv6 RIP interface with prefixes defined:

```
IPV6_RIP_INTERFACE
    NAME=OSAQDIO3L6
    PREFIX=2001:0DB8:3::/48
    PREFIX=2001:0DB8:4::/48;
```

  The prefixes defined in this manner on an IPv6 RIP interface are advertised into the IPv6 RIP autonomous system as reachable if IPv6 RIP filters permit it. They are also advertised into the IPv6

OSPF autonomous system as reachable, if OMPROUTE is running IPv6 OSPF and is configured as an IPv6 AS boundary router importing RIP routes.

The following sample shows an IPv6 generic interface with prefixes defined:

```
IPV6_INTERFACE
    NAME=OSAQDIO2L6
    PREFIX=2001:0DB8:5::/48
    PREFIX=2001:0DB8:6::/48;
```

The prefixes defined in this manner on an IPv6 generic interface are advertised into the RIP autonomous system as reachable, if OMPROUTE is running IPv6 RIP and IPv6 RIP filters permit it. They are also advertised into the IPv6 OSPF autonomous system as reachable, if OMPROUTE is running IPv6 OSPF and is configured as an IPv6 AS boundary router importing direct routes.

Method of assigning interface definitions to stack interfaces (wildcard and explicit):

For IPv6 interfaces, interface-name wildcards can be used to simplify definitions. However, be sure to understand how they are parsed, and how different types of interface definitions interact with each other, to avoid unintended results. The following outline shows the algorithm OMPROUTE uses to find the matching definitions in the OMPROUTE configuration file for an IPv6 stack interface.

a. Search for an IPv6_RIP_Interface definition for the interface as follows:
   i) Search for an explicit matching IPv6_RIP_Interface statement for the interface. This match is one where the name parameter exactly matches the interface name. If one is found, use that definition and go to step b.
   ii) Search for the best IPv6_RIP_Interface wildcard match for the name. The IPv6_RIP_Interface wildcard definitions are searched, starting with the most specific (longest wildcard name string) and checking each in order of declining specificity until a match is found. As soon as a match is found, use that definition and go to step b.
b. Search for an IPv6_OSPF_Interface definition for the interface. This step is done regardless of the outcome of step a. The steps for searching IPv6_OSPF_Interface definitions are the same as the steps for searching IPv6_RIP_Interface definitions, except that IPv6_OSPF_Interface definitions are searched.
c. If either an IPv6_RIP_Interface or an IPv6_OSPF_Interface definition, or both, are found, the algorithm is complete. In this case, IPv6_Interface definitions are not searched. If neither an IPv6_RIP_Interface or an IPv6_OSPF_Interface definition was found, go to step d.
d. Search for an IPv6_Interface definition for the interface. The steps for searching IPv6_Interface definitions are the same as the steps for searching IPv6_RIP_Interface statements, except that IPv6_Interface definitions are searched.
e. If no definitions are found, check the value of Global_Options Ignore_Undefined_Interfaces. If this option is turned on, the interface is ignored. If it is not turned on, the interface is treated as if it were defined with an IPv6_Interface statement. Default values are used for all parameters.

The algorithm is complete. Key conclusions of this algorithm are as follows:

- If an IPv6_RIP_Interface definition, an IPv6_OSPF_Interface definition, or both, are found, IPv6_Interface definitions are not considered. This means that any matching IPv6_RIP_Interface or IPv6_OSPF_Interface definition supersedes all IPv6_Interface definitions, even if the IPv6_Interface definitions are explicit or more specific wildcard matches. For example, an IPv6_OSPF_Interface definition with a name parameter of V* supersedes any IPv6_Interface, explicit or wildcard, with a name parameter that begins with V. In this case, the IPv6_Interface definition is redundant and can never be used. If OMPROUTE detects this case, it issues message EZZ8068I and deletes the redundant IPv6_Interface definition.

  Note: If an IPv6_Interface definition was already selected for an interface that is installed in the stack, and then an IPv6_OSPF_Interface or IPv6_RIP_Interface definition that would make that IPv6_Interface definition redundant is added by using RECONFIG, OMPROUTE issues message EZZ8069I and retains the IPv6_Interface definition.

- An interface can be both an IPv6_RIP_Interface and an IPv6_OSPF_Interface. OMPROUTE supports running both protocols over the same interface. However, an interface cannot be both an interface that runs no routing protocol (that is, defined with the IPv6_Interface statement) and one that runs IPv6 RIP, IPv6 OSPF, or both.

7. Define interface costs (OSPF_INTERFACE, RIP_INTERFACE, IPV6_OSPF_INTERFACE, and IPV6_RIP_INTERFACE).

   Both the OSPF and RIP protocols have a cost value that is associated with interfaces. With both protocols, the cost of a route to reach a destination is the sum of the costs of each link to be traversed on the way to the destination. In the sample network that is shown in Figure 48 on page 282, the cost of a route to get from TCPCS7 to router 3.3.3.3 through TCPCS4 is the cost of the link from TCPCS7 to TCPCS4 plus the cost of the link from TCPCS4 to router 3.3.3.3.

   The method for configuring cost values differs between the OSPF and RIP protocols. Configure the cost values of OSPF links to ensure that preferred routes to destinations have a lower cost than less preferable routes. The less preferable routes, with the higher cost, are not used except upon failure of the preferred routes.

   For the following examples, the sample network that is shown in Figure 48 on page 282 is used and the convention stack (interface) is used to refer to the cost configured for a particular interface on a stack. For instance, TCPCS7(9.67.106.7) refers to the cost configured for interface 9.67.106.7 on TCPCS7. While these examples use the IPv4 portion of the sample network, the same methods for computing route costs would also be used by the IPv6 portion.

   The following three routes are possible from TCPCS7 to router 3.3.3.3:

   - Direct (TCPCS7 to 3.3.3.3)
   - Through TCPCS4 (TCPCS7 to TCPCS4 to 3.3.3.3)
   - Through router 8.8.8.8 and TCPCS4 (TCPCS7 to 8.8.8.8 to TCPCS4 to 3.3.3.3)

   If the preferred route from TCPCS7 to router 3.3.3.3 is through TCPCS4, then interface costs must be configured such that the following conditions are true:

   ```
   TCPCS7(9.67.106.7) + TCPCS4(9.67.101.4) < TCPCS7(9.67.102.7)
   TCPCS7(9.67.106.7) + TCPCS4(9.67.101.4) < TCPCS7(9.67.100.7) +
   8.8.8.8(9.67.105.8) + TCPCS4(9.67.101.4)
   ```

   The reasons for preferring one route over another are numerous. One approach for assigning OSPF link costs would be to set the costs to values inversely proportional to the bandwidth of the physical media. This would result in higher bandwidth routes with lower costs, thus becoming the preferred routes.

   The cost values of RIP links are generally set to a value of 1. This results in the cost of a route to a destination being the number of hops to reach the destination. In the sample network, this would result in the three possible RIP routes from TCPCS7 to router 3.3.3.3 having the following costs:

   - Direct (TCPCS7 to 3.3.3.3), cost = 1
   - Through TCPCS4 (TCPCS7 to TCPCS4 to 3.3.3.3), cost = 2
   - Through router 8.8.8.8 and TCPCS4 (TCPCS7 to 8.8.8.8 to TCPCS4 to 3.3.3.3), cost = 3

   If it was desirable that the route through TCPCS4 be the preferred route, this can be accomplished by increasing the cost of getting directly from TCPCS7 to router 3.3.3.3. This can be done by increasing either the out metric for 9.67.102.3 on router 3.3.3.3 or the in metric for 9.67.102.7 on TCPCS7. Take care when you are increasing in metric and out metric values to be sure that the cost to reach any destination does not exceed the RIP maximum of 15.

   - IPv4 OSPF and RIP

     The cost value of an OSPF interface is set with the COST0 parameter of the OSPF_INTERFACE statement. The in metric and out metric of a RIP interface are set with the IN_METRIC and OUT_METRIC parameters of the RIP_INTERFACE statement.
   - IPv6 OSPF and RIP

The cost value of an IPv6 OSPF interface is set with the COST parameter of the IPV6_OSPF_INTERFACE statement. The in metric and out metric of an IPv6 RIP interface are set with the IN_METRIC and OUT_METRIC parameters of the IPV6_RIP_INTERFACE statement.

8. Configure virtual links, if the OSPF protocol is used.

   The OSPF protocol is dependent upon complete connectivity of the backbone area. To maintain backbone connectivity, each backbone router must be interconnected. If the configuration of an OSPF autonomous system is such that the backbone area becomes separated into two or more disconnected sections, connectivity must be restored for the protocol to work correctly. This can be done by using a virtual link. Do not confuse an OSPF virtual link with a VIPA link. Virtual links can be configured between any two backbone routers that have an interface to a common non-backbone area.

   - IPv4 OSPF

     The VIRTUAL_LINK statements specify the router ID of the link endpoint and must be configured at both endpoints. In the sample network that is shown in Figure 48 on page 282, a virtual link is configured between TCPCS4 and TCPCS7 to restore backbone connectivity through area 1.1.1.1.

     TCPCS4:

     ```
     VIRTUAL_LINK
        Virtual_Endpoint_RouterID=7.7.7.7
        Links_Transit_Area=1.1.1.1;
     ```

     TCPCS7:

     ```
     VIRTUAL_LINK
        Virtual_Endpoint_RouterID=4.4.4.4
        Links_Transit_Area=1.1.1.1;
     ```

   - IPv6 OSPF

     The IPV6_VIRTUAL_LINK statements specify the router ID of the link endpoint and must be configured at both endpoints. In the sample network, a virtual link is configured between TCPCS64 and TCPCS67 to restore backbone connectivity through area 6.6.6.6.

     TCPCS64:

     ```
     IPV6_VIRTUAL_LINK
        Virtual_Endpoint_RouterID=67.67.67.67
        Links_Transit_Area=6.6.6.6;
     ```

     TCPCS67:

     ```
     IPV6_VIRTUAL_LINK
        Virtual_Endpoint_RouterID=64.64.64.64
        Links_Transit_Area=6.6.6.6;
     ```

9. Manage high-cost links, if the OSPF protocol is used.

   The periodic nature of OSPF routing traffic requires a link's underlying data-link connection to be constantly open. This can result in unwanted usage charges on network segments whose costs are high. There are two configuration steps that can be taken to inhibit the periodic nature of the protocol.

   The first step that can be taken is to define the link as a demand circuit. When this is done, link state advertisements (LSAs) sent over the interface are not periodically refreshed. Only LSAs with real changes are readvertised. In addition, aging of these LSAs is disabled such that they will not age out of the link state database.

Another step that can be taken is to define hello suppression for the link. Hello suppression is meaningful only if the link is a demand circuit and is either point-to-point or point-to-multipoint. Hello suppression inhibits the periodic transmission of OSPF hello packets.

- IPv4 OSPF

  To define OSPF interfaces as demand circuits, the Demand_Circuit=YES parameter must first be specified on the global OSPF configuration statement. Then, the OSPF_INTERFACE statement for each interface to be configured as a demand circuit must be specified with the Demand_Circuit=YES parameter. Use the Hello_Suppression parameter of the OSPF_INTERFACE statement to configure hello suppression. For more information about configuring the Hello_Suppression parameter on the OSPF_INTERFACE statement, see z/OS Communications Server: IP Configuration Reference. If hello suppression is implemented, the PP_Poll_Interval parameter of the OSPF_INTERFACE statement can be used to specify the interval at which OMPROUTE attempts to contact a neighbor to reestablish a neighbor relationship when the relationship failed, but the interface is still available.

- IPv6 OSPF

  To define IPv6 OSPF interfaces as demand circuits, the Demand_Circuit=YES parameter must first be specified on the global IPV6_OSPF configuration statement. Then, the IPV6_OSPF_INTERFACE statement for each interface to be configured as a demand circuit must be specified with the Demand_Circuit=YES parameter. Use the Hello_Suppression parameter of the IPV6_OSPF_INTERFACE statement to configure hello suppression. For more information about configuring the Hello_Suppression parameter on the IPV6_OSPF_INTERFACE statement, see z/OS Communications Server: IP Configuration Reference. If hello suppression is implemented, the PP_Poll_Interval parameter of the IPV6_OSPF_INTERFACE statement can be used to specify the interval at which OMPROUTE attempts to contact a neighbor to reestablish a neighbor relationship when the relationship failed, but the interface is still available.

10. Define RIP filters, if the RIP protocol is used.

    RIP Filters can be configured to OMPROUTE such that certain RIP routing information is not broadcast out to other routers and/or accepted from other routers. The filters can be applied to individual RIP interfaces or to all RIP interfaces. When you are defining a filter, a filter type (sending or receiving) is specified along with values that identify the route information to be filtered. By using filters, an installation can limit the amount of RIP routing information broadcast into the network and/or the amount of RIP routing information that is maintained by OMPROUTE. In addition, filters can be used to hide destination addresses from portions of the network.

    - IPv4 RIP

      To configure a filter for an individual RIP interface, use the FILTER parameter of the RIP_INTERFACE statement. To configure a filter that applies to all IPv4 RIP interfaces, use the global FILTER statement. In the sample network that is shown in Figure 48 on page 282, if you wanted to hide the 10.1.1.0 subnet from TCPCS6 (and all routers and hosts on the remote side of TCPCS6), you can define the following filter on TCPCS7:

      ```
      Filter=(nosend,10.1.1.0,255.255.255.0);
      ```

    - IPv6 RIP

      To configure a filter for an individual IPv6 RIP interface, use the FILTER parameter of the IPV6_RIP_INTERFACE statement. To configure a filter that applies to all IPv6 RIP interfaces, use the global IPV6_RIP_FILTER statement. In the sample network that is shown in Figure 48 on page 282, for example, if you wanted to hide the 2001:0DB8:0:A1B/64 prefix from TCPCS4, you can define the following filter on TCPCS4:

      ```
      IPv6_RIP_Filter=(noreceive,2001:0DB8:0:A1B/64);
      ```

11. Define route precedence in a multiprotocol environment, if the OSPF protocol is used.

    This discussion of route precedence is complicated. If only the OSPF or IPv6 OSPF routing protocol, or both, are used in your network, route precedence is less of a concern. If, in addition, none of your OSPF or IPv6 OSPF routers are configured as AS boundary routers, the route precedence concern is

entirely eliminated. For environments with multiple protocols or AS boundary routers, the following information is provided. In this discussion, RIP is meant to apply to both RIP and IPv6 RIP, OSPF is meant to apply to both OSPF and IPv6 OSPF, and the OSPF configuration statement is meant to apply to both the OSPF statement and the IPV6_OSPF statement.

OMPROUTE applies an order of precedence in choosing between two routes to the same destination that were learned through different routing protocols or by using information that is provided by an OSPF AS boundary router. To describe this order of precedence that is applied by OMPROUTE, a few terms must first be defined.

RIP route
A route learned through the RIP protocol. A RIP route is generated by using information that is provided in a RIP packet from a neighboring router. For example, in the sample network that is shown in Figure 48 on page 282, the route from TCPCS7 to destination subnet 30.1.1.0 is a RIP route.

OSPF internal route
A route learned through the OSPF protocol where the entire path traversed to reach the destination lies within the OSPF autonomous system. For example, in the sample network that is shown in Figure 48 on page 282, the route from TCPCS7 to destination 9.67.108.2 on Router 2.2.2.2 is an OSPF internal route.

OSPF external route
A route learned through the OSPF protocol where part of the path traversed to reach the destination does not lie within the OSPF autonomous system. The path leaves the autonomous system when it uses information brought into the OSPF autonomous system by an AS boundary router. This information that is brought into the OSPF AS can be information that is imported from a different autonomous system (for example, RIP) or information about destinations that are statically configured on or directly connected to the AS boundary router. For example, in the sample network that is shown in Figure 48 on page 282, the route from TCPCS4 to destination 9.67.103.6 on TCPCS6 is an OSPF external route. TCPCS7, configured as an AS boundary router, imported information about that destination into the OSPF AS from the RIP AS.

OSPF external routes fall into two categories that are based upon the setting of the multiprotocol comparison value. If the comparison value is set to Type1 on the AS boundary router that imports the external information into the OSPF AS, then OSPF external routes that are generated by using this information are OSPF type 1 external routes. If the comparison value is set to Type2 on the AS boundary router, then the generated routes are OSPF type 2 external routes. For example, in the sample network that is shown in Figure 48 on page 282, if the comparison value on TCPCS7 (an AS boundary router) is set to Type1, the route from TCPCS4 to destination 9.67.103.6 on TCPCS6 is an OSPF type 1 external route. If the comparison value on TCPCS7 is set to Type2, the route is an OSPF type 2 external route.

Multiprotocol comparison

You can configure this comparison value to allow for the specification of how route costs from different autonomous systems are treated when they coexist. In OMPROUTE, you can configure this value by using the COMPARISON parameter on the OSPF or IPV6_OSPF configuration statements. When COMPARISON=Type1 is configured, the route cost values used within different autonomous systems (for example, the OSPF AS and the RIP AS) are considered comparable. With COMPARISON=Type2 configured, the route cost values used with the different autonomous systems are considered non-comparable.

The comparison value can be used in several different ways, depending on the function that is being provided by a router:

- As an AS boundary router, OMPROUTE uses the comparison value to determine the type of external routes (type 1 or type 2) that are generated by routers in the OSPF AS by using routing information that the AS boundary router imports into the OSPF AS.
- As an AS boundary router, OMPROUTE also uses the comparison value in determining how route cost values are assigned when routes are imported from the OSPF AS into the RIP AS.

– When COMPARISON=Type1 is configured (indicating that cost values are comparable), an OSPF route that is imported into the RIP AS is advertised with the actual cost of the OSPF route. For example, in the sample network, if TCPCS7 is configured with COMPARISON=Type1 and the OSPF route from TCPCS7 to destination 9.67.108.2 on TCPCS2 has a cost of 7, then TCPCS7 advertises into the RIP AS a RIP route to that destination with a cost of 7.

  Notes:

  a. An exception to this rule (defining how OSPF routes are advertised into the RIP AS when COMPARISON=Type1) occurs when the OSPF route to be imported is an OSPF type 2 external route. In this case, the route is not advertised into the RIP AS at all.

  b. It is important to remember the requirement that all destinations in the RIP AS must be reachable with a cost no greater than 15. Using COMPARISON=Type1 requires that the cost values of OSPF routes be low. Any destinations in the OSPF AS that can be reached only from the RIP AS with a cost greater than 15 become unreachable.

– When COMPARISON=Type2 is configured (indicating that cost values are non-comparable), an OSPF route that is imported into the RIP AS is advertised with a cost of 1. If a router in the RIP AS has two possible routes to a destination, one internal to the RIP AS and another that was imported from OSPF, this approach results in the route that is imported from OSPF being favored. For example, in the sample network that is shown in Figure 48 on page 282, if TCPCS7 is configured with COMPARISON=Type2 and TCPCS7 can somehow reach a destination in the 30.1.1.0 subnet without passing through TCPCS6 (by using links that are not shown in the sample), then TCPCS7 advertises into the RIP AS a RIP route to the destination with a cost of 1. As a result, TCPCS6 determines that the destination can be reached through TCPCS7 with a cost of 2. If the cost of the route for TCPCS6 to reach the destination internal to the RIP AS is greater than 2, then the route through TCPCS7 is chosen.

  Note: An exception to this rule (defining how OSPF routes are advertised into the RIP AS when COMPARISON=Type2) occurs when the OSPF route to be imported is an OSPF type 2 external route. In this case, the route is advertised into the RIP AS with the actual cost of the OSPF type 2 external route.

- As any router that has routing information from different autonomous systems, OMPROUTE uses the comparison value while it is choosing between the routes generated by using the information from the different autonomous systems. How the comparison value is used in this case is shown in Table 19 on page 323.

Given these definitions, the order of precedence that is used in choosing between multiple routes to the same destination, which were learned through the different protocols or by using information that is provided by an OSPF AS boundary router, can be shown in Table 19 on page 323. In Table 19 on page 323, *Source comparison* refers to the setting of the comparison value (by using the COMPARISON parameter on the OSPF configuration statement) on the router that is using the order of precedence to choose between the multiple routes. Route 1 and Route 2 are the two possible routes that are being chosen between.

Table 19. Route precedence

Source comparison	Route 1 type	Route 2 type	Route chosen
Type 1	OSPF internal	RIP	OSPF internal
Type 1	OSPF internal	OSPF type 1 external	OSPF internal
Type 1	OSPF internal	OSPF type 2 external	OSPF internal
Type 1	RIP	OSPF type 1 external	Lowest cost route
Type 1	RIP	OSPF type 2 external	RIP route
Type 1	OSPF type 1 external	OSPF type 2 external	OSPF type 1 external
Type 2	OSPF internal	RIP	OSPF internal

Table 19. Route precedence (continued)			
Source comparison	**Route 1 type**	**Route 2 type**	**Route chosen**
Type 2	OSPF internal	OSPF type 1 external	OSPF internal
Type 2	OSPF internal	OSPF type 2 external	OSPF internal
Type 2	RIP	OSPF type 1 external	OSPF type 1 external
Type 2	RIP	OSPF type 2 external	Lowest cost route
Type 2	OSPF type 1 external	OSPF type 2 external	OSPF type 1 external

Minimizing the routing responsibility of z/OS Communications Server

OMPROUTE may be run on z/OS Communications Server for a variety of reasons. If the z/OS Communications Server host is being used as an application or server host and the routing daemon is being run primarily to provide access to network resources, or to provide network resources access to the z/OS Communications Server host, then care must be taken to ensure that the z/OS Communications Server host is not overly burdened with routing work. Unlike routers or other network boxes whose sole purpose is routing, an application host z/OS Communications Server will be doing many things other than routing, and it is not desirable for a large percentage of machine resources (memory and CPU) to be used for routing tasks, as can happen in very complex or unstable networks. In this case the z/OS Communications Server should not be configured as a backbone router, either intentionally or inadvertently. Careful network design can minimize the routing burdens on the z/OS Communications Server application host without compromising the accessibility of z/OS Communications Server resources to the network and vice versa. If care is not taken to minimize the routing work required by the z/OS Communications Server host, OMPROUTE may consume excessive cycles or memory processing huge numbers of routing updates from the network. Or the burden of routing updates may become so large that the z/OS Communications Server cannot keep up because of other workloads on the machine. Because OSPF is heavily timer-driven, this could cause loss of adjacencies and routing problems.

The primary way to reduce the routing burdens on the z/OS Communications Server host is by use of OSPF areas. See step “3” on page 306 for more information. A z/OS Communications Server application host or sysplex can be placed into a non-backbone area with dedicated routers acting as area-border routers. The area-border routers would advertise the z/OS Communications Server resources to other attached areas (for example the backbone) and would summarize the network outside the local area to the z/OS Communications Server hosts. If possible, this can be further refined to reduce routing protocol traffic by use of interarea route summarization, accomplished in OMPROUTE area-border routers by the RANGE and IPV6_RANGE statements, and in Cisco routers with the area range command. For more information, see z/OS Communications Server: IP Configuration Reference and step “4” on page 307.

An even further, and ideal, optimization would be to make the area containing the z/OS Communications Server application host or sysplex a stub area. A stub area can be configured such that route summaries (IPv4) or prefixes (IPv6) from other areas are not flooded into the stub area by the area border routers. When this is done, only routes to destinations within the stub area are shared among the hosts. Default routes are used to represent all destinations outside the stub area. The stub area's resources are still advertised to the network at large by the area-border routers. You can use this optimization, sometimes referred to as a totally stubby area, if the following conditions apply to your network:

- It is acceptable to use default routes to reach destinations outside the stub area. This means that either there is only one area-border router connecting the stub area to the rest of the network, or if there are multiple such connections they are redundant, so that it does not matter which one is used to get outside the stub area.
- You have no non-OSPF destinations to advertise to the network at large. Stub areas do not permit importation of OSPF external routes. This means for example that you do not have a RIP network attached to the stub area, or if you do, you do not want its destinations reachable from the stub area. Other types of routes that cannot be imported into stub areas include direct routes (for example,

for networks attached to interfaces that are not running the OSPF protocol) and static routes. RIP or static routes can be used by the z/OS Communications Server that learns them, they just cannot be advertised.

Tip: If you define your VIPAs as OSPF interfaces in your OMPROUTE configuration file, routes to their addresses will be considered OSPF routes, and therefore importable into the stub area and able to be advertised by the area-border routers to the network at large.

It is highly recommended to put z/OS Communications Server application hosts or sysplexes into stub areas if at all possible.

A further optimization is to prevent z/OS Communications Server from becoming the designated router on multiaccess media, when pure routers that can perform this function are present. On a multiaccess medium, the designated router and the backup designated router will carry the majority of the routing protocol load for all hosts on the medium. While z/OS Communications Server is capable of performing this role, it does impose additional routing overhead on the system. It would be preferable to allow pure routers to perform this role, if they are available. This is accomplished by ensuring that the pure routers' interfaces onto the medium have higher ROUTER_PRIORITY values than the z/OS Communications Server interfaces on the same medium. However, if the only hosts on a medium are z/OS Communications Server, such as in HiperSockets communications, then one or two of them will have to be designated router or backup designated router.

Tip: You can use the IBM Health Checker for z/OS to check whether the total number of indirect routes in a TCP/IP stack routing table exceeds a maximum threshold. When this threshold is exceeded, OMPROUTE and the TCP/IP stack can experience high CPU consumption from routing changes. For more information about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide and IBM Health Checker for z/OS: User's Guide.

Preventing futile neighbor state loops during adjacency formation

In OSPF environments in which there might be a problem with some remote hardware (for example, router, switch, or network cable) that is beyond detection by z/OS hardware or software, OMPROUTE can get into an infinite (or futile) neighbor state loop over one of its interfaces with a neighbor. This loop might contribute to increased workload. In LAN configurations in which there are parallel OSPF interfaces that can reach the same neighbor for adjacency formation, unless you are using OMPROUTE futile neighbor state loop detection or unless you manually fix the problem, the backup interfaces are not used until after an outage occurs for the OSPF interface that was initially involved in an adjacency formation attempt with a designated router. For information about parallel OSPF interfaces, see step “5” on page 309.

Futile neighbor state loops can occur during adjacency formation with a neighboring designated router before full adjacency has been reached. According to OSPF protocol, all routers on a multiaccess network attempt to establish adjacency with the designated router (DR) and backup designated router (BDR) after two-way communication has been established. When a problem occurs during the adjacency formation process, the adjacency formation attempt is restarted. Repeated adjacency attempts can continue until full adjacency has been established or an interruption occurs. Interruption occurs from an interface outage to the neighbor or from a neighbor outage. Whether or not parallel interfaces are used, OMPROUTE repeatedly attempts to establish full adjacency with the neighbor, and if the problem is due to a cabling error or failure or some other non-transient condition, it is highly unlikely that the problem will be resolved and the loop will become futile.

To stop a futile neighbor state loop, you can either manually deactivate the interface or suspend the OSPF interface within OMPROUTE so that OMPROUTE will stop attempting to form the adjacency over the interface. If the OSPF interface is suspended, any active sessions using static routes over the interface will not be disrupted. If the interface is deactivated, all active sessions over the interface will be disrupted. If a parallel OSPF interface is available, OMPROUTE will then attempt to form adjacency with the neighbor over the parallel interface.

OMPROUTE futile neighbor state loop detection removes the need for manual detection of futile neighbor state loops, manual intervention to resolve the loops, and the disruption of existing sessions due to deactivating the interface. Code the DR_MAX_ADJ_ATTEMPT parameter on the OSPF or IPV6_OSPF statement, or both, in your OMPROUTE configuration file to enable this function. OMPROUTE will then

report and control futile neighbor state loops during the adjacency formation process. For a list of interfaces that support the futile neighbor state loop detection function, see z/OS Communications Server: IP Configuration Reference

If you use the DR_MAX_ADJ_ATTEMPT parameter, futile neighbor state loops are automatically detected and reported using message EZZ8157I. If a parallel OSPF interface is not available, adjacency formation attempts continue to be tried over the same interface. If parallel OSPF interfaces are available, an interface change is reported using message EZZ8158I and adjacency formation attempts are tried again over a parallel OSPF interface. The problematic interface is suspended within OMPROUTE, but the interface is not deactivated and active sessions over the interface are not disrupted.

After a problematic OSPF interface is suspended in OMPROUTE and adjacencies are formed on a parallel OSPF interface, you might want to switch back to the original interface after you have fixed the problem that caused the futile neighbor state loop. To accomplish this, activate the repaired interface and then suspend the parallel interface. After OSPF adjacencies are established over the repaired interface, the parallel interface can be reactivated so it is again available as a backup for the repaired interface. Use the ACTIVATE function of the OMPROUTE MODIFY command to activate a suspended OSPF interface. Use the SUSPEND function of the OMPROUTE MODIFY command to manually suspend an OSPF interface. For more information about these functions of the MODIFY command for OMPROUTE, see z/OS Communications Server: IP System Administrator's Commands.

Verification of OMPROUTE IPv4 configuration and state

The following topics show sample output from each of the commands that can be used to display OMPROUTE IPv4 information. The syntax of these DISPLAY commands, as well as detailed information about the data displayed, can be found in z/OS Communications Server: IP System Administrator's Commands.

Note: All commands that include the LIST subparameter indicate that the information being displayed is configured information only and does not necessarily mean that the information is currently being used by OMPROUTE. To display information in current use, use related commands to display current, run-time statistics, and parameters. There are cases when the configured information will not match the in-use information due to some undefined or unresolved information in the OMPROUTE configuration. For example, undefined interfaces or parameters in the OMPROUTE configuration or an incorrect sequence of dynamic reconfiguration with the MODIFY OMPROUTE,RECONFIG command can result in no update of the in-use information at all. Information defined on wildcard interfaces is not displayed in the LIST commands; it is displayed only in the corresponding non-LIST commands when wildcard information is resolved to actual physical interfaces.

Displaying all OSPF configuration information

To display all of the OSPF configuration information, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,ALL
EZZ7831I GLOBAL CONFIGURATION 735
    TRACE: 0, DEBUG: 0, SADEBUG LEVEL: 0
    STACK AFFINITY:          TCPCS7
    OSPF PROTOCOL:           ENABLED
    EXTERNAL COMPARISON:     TYPE 2
    AS BOUNDARY CAPABILITY:  ENABLED
    IMPORT EXTERNAL ROUTES:  RIP SUB
    ORIG. DEFAULT ROUTE:     NO
    DEFAULT ROUTE COST:      (1, TYPE 2)
    DEFAULT FORWARD. ADDR.:  0.0.0.0
    LEARN HIGHER COST DFLT:  NO
    DEMAND CIRCUITS:         ENABLED

EZZ7832I AREA CONFIGURATION
AREA ID          AUTYPE          STUB? DEFAULT-COST IMPORT-SUMMARIES?
0.0.0.0          0=NONE           NO          N/A           N/A
1.1.1.1          0=NONE           NO          N/A           N/A

--AREA RANGES--
AREA ID          ADDRESS          MASK             ADVERTISE?
1.1.1.1          9.67.101.0       255.255.255.0    NO
```

```
EZZ7833I INTERFACE CONFIGURATION
IP ADDRESS      AREA            COST  RTRNS  TRNSDLY PRI  HELLO   DEAD    DB_EX
7.7.7.7         1.1.1.1            1    N/A    N/A   N/A    N/A    N/A     N/A
9.67.104.7      1.1.1.1            1      5      1     1     10     40      40
9.67.100.7      1.1.1.1            1      5      1     1     10     40      40
9.67.102.7      1.1.1.1            1      5      1     1     10     40      40
9.67.106.7      1.1.1.1            1      5      1     1     10     40      40
9.67.107.7      0.0.0.0            1      5      1     1     10     40      40

EZZ7836I VIRTUAL LINK CONFIGURATION
VIRTUAL ENDPOINT     TRANSIT AREA      RTRNS  TRNSDLY HELLO  DEAD  DB_EX
4.4.4.4              1.1.1.1             10      5     30   180    180

EZZ7835I NBMA CONFIGURATION
                INTERFACE ADDR      POLL INTERVAL
                     9.67.104.7          180
EZZ7834I NEIGHBOR CONFIGURATION
                NEIGHBOR ADDR     INTERFACE ADDRESS   DR ELIGIBLE?
                9.67.104.15       9.67.104.7          YES
                9.67.104.25       9.67.104.7          NO
                9.67.104.16       9.67.104.7          NO
```

Displaying information about configured OSPF areas

To display information about configured OSPF Areas, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,AREAS
EZZ7832I AREA CONFIGURATION 737
AREA ID          AUTYPE          STUB? DEFAULT-COST IMPORT-SUMMARIES?
0.0.0.0          0=NONE           NO          N/A           N/A
1.1.1.1          0=NONE           NO          N/A           N/A

--AREA RANGES--
AREA ID          ADDRESS          MASK             ADVERTISE?
1.1.1.1          9.67.101.0       255.255.255.0    NO
```

Displaying configuration information about configured OSPF interfaces

To display configuration information about configured OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,IFS
EZZ7833I INTERFACE CONFIGURATION 739
IP ADDRESS      AREA            COST  RTRNS  TRNSDLY PRI  HELLO   DEAD   DB_EX
7.7.7.7         1.1.1.1            1    N/A    N/A   N/A    N/A    N/A    N/A
9.67.104.7      1.1.1.1            1      5      1     1     10     40     40
9.67.100.7      1.1.1.1            1      5      1     1     10     40     40
9.67.102.7      1.1.1.1            1      5      1     1     10     40     40
9.67.106.7      1.1.1.1            1      5      1     1     10     40     40
9.67.107.7      0.0.0.0            1      5      1     1     10     40     40
```

Note: Wildcard interface definitions are not displayed. However, when an actual interface is resolved to a wildcard definition, its information is displayed.

Displaying information about configured Non-broadcast Multiple Access OSPF interfaces

To display information about configured Non-broadcast Multiple Access OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,NBMA
EZZ7835I NBMA CONFIGURATION 745
                INTERFACE ADDR      POLL INTERVAL
                9.67.104.7          180
```

Displaying information about configured OSPF virtual links

To display information about configured OSPF virtual links, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,VLINKS
EZZ7836I VIRTUAL LINK CONFIGURATION 747
```

```
VIRTUAL ENDPOINT     TRANSIT AREA      RTRNS  TRNSDLY HELLO  DEAD  DB_EX
4.4.4.4              1.1.1.1              10      5     30   180    180
```

Displaying information about configured OSPF neighbors

To display information about configured OSPF neighbors enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LIST,NBRS
EZZ7834I NEIGHBOR CONFIGURATION 749
                NEIGHBOR ADDR     INTERFACE ADDRESS   DR ELIGIBLE?
                9.67.104.15       9.67.104.7          YES
                9.67.104.25       9.67.104.7          NO
                9.67.104.16       9.67.104.7          NO
```

Displaying the contents of a single OSPF link state advertisement

To display the contents of a single OSPF link state advertisement, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,LSA,LSTYPE=1,LSID=7.7.7.7,ORIG=7.7.7.7,AREAID=1.1.1.1
EZZ7880I LSA DETAILS 751
        LS AGE:          521
        LS OPTIONS:      E,DC
        LS TYPE:         1
        LS DESTINATION (ID): 7.7.7.7
        LS ORIGINATOR:   7.7.7.7
        LS SEQUENCE NO:  0X80000013
        LS CHECKSUM:     0XA9A
        LS LENGTH:       120
        ROUTER TYPE:  ABR,ASBR,V
        # ROUTER IFCS:   8
                LINK ID:          7.7.7.4
                LINK DATA:        255.255.255.252
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0  METRIC:  1
                LINK ID:          8.8.8.8
                LINK DATA:        9.67.100.7
                INTERFACE TYPE:   1
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1 (1)
                LINK ID:          3.3.3.3
                LINK DATA:        9.67.102.7
                INTERFACE TYPE:   1
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1 (1)
                LINK ID:          4.4.4.4
                LINK DATA:        9.67.106.7
                INTERFACE TYPE:   1
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1 (1)
                LINK ID:          7.7.7.7
                LINK DATA:        255.255.255.255
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1
                LINK ID:          9.67.100.8
                LINK DATA:        255.255.255.255
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1
                LINK ID:          9.67.102.3
                LINK DATA:        255.255.255.255
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1
                LINK ID:          9.67.106.4
                LINK DATA:        255.255.255.255
                INTERFACE TYPE:   3
                        NO. OF METRICS: 0
                        TOS 0 METRIC:   1
```

Displaying statistics and parameters for OSPF areas

To display statistics and parameters for all OSPF areas attached to the router, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,AREASUM
EZZ7848I AREA SUMMARY 757
AREA ID        AUTHENTICATION   #IFCS  #NETS  #RTRS  #BRDRS DEMAND
0.0.0.0           NONE             2      0      4      2  ON
1.1.1.1           NONE             5      0      4      2  ON
```

Displaying the list of AS external advertisements

To display a list of AS external advertisements that are in the OSPF link state database, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,EXTERNAL
EZZ7853I AREA LINK STATE DATABASE 759
TYPE LS DESTINATION     LS ORIGINATOR     SEQNO     AGE   XSUM
  5 @6.6.6.6            7.7.7.7         0X80000007  825  0X1B5C
  5 @9.67.103.6         7.7.7.7         0X80000007  831  0XE1F3
  5 @10.1.1.0           2.2.2.2         0X80000003 1690  0X2775
  5 @10.1.1.1           2.2.2.2         0X80000003 1690  0X1D7E
  5 @20.1.1.0           5.5.5.5         0X80000003 1616  0X4A3C
  5 @20.1.1.1           5.5.5.5         0X80000003 1616  0X4045
  5 @30.0.0.0           7.7.7.7         0X80000006  831  0XB0C0
  5 @30.1.1.0           7.7.7.7         0X80000006  831  0X99D5
  5 @30.1.1.4           7.7.7.7         0X80000001  825  0X7BF4
  5 @30.1.1.8           7.7.7.7         0X80000001  825  0X5319
  5 @130.200.0.0        3.3.3.3         0X80000003 1695  0X98C0
  5 @130.200.0.0        8.8.8.8         0X80000003 1630  0X243
  5 @130.200.1.1        3.3.3.3         0X80000003 1695  0X83D3
  5 @130.200.1.18       8.8.8.8         0X80000003 1630  0X42EF
  5 @130.201.0.0        3.3.3.3         0X80000003 1695  0X8CCB
  5 @130.201.0.0        8.8.8.8         0X80000003 1630  0XF54E
  5 @130.202.0.0        3.3.3.3         0X80000003 1694  0X80D6
  5 @130.202.0.0        8.8.8.8         0X80000003 1629  0XE959
               # ADVERTISEMENTS:       18
               CHECKSUM TOTAL:         0X83472
```

Displaying a list of non-AS external advertisements

To display a list of non-AS external advertisements that are in the OSPF link state database for a particular OSPF area, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,DATABASE,AREAID=1.1.1.1
EZZ7853I AREA LINK STATE DATABASE 761
TYPE LS DESTINATION     LS ORIGINATOR     SEQNO     AGE   XSUM
  1 @3.3.3.3            3.3.3.3         0X8000000F  879  0X8B11
  1 @4.4.4.4            4.4.4.4         0X8000001A  713  0XA020
  1 @7.7.7.7            7.7.7.7         0X80000013  711  0XA9A
  1 @8.8.8.8            8.8.8.8         0X8000000D  861  0XBD81
  3 @2.2.2.2            4.4.4.4         0X80000003 1676  0XC45C
  3 @5.5.5.4            7.7.7.7         0X80000003  880  0XE327
  3 @5.5.5.5            7.7.7.7         0X80000003  880  0XDF29
  3 @7.7.7.4            7.7.7.7         0X80000001  710  0X956E
  3 @9.67.107.5         7.7.7.7         0X80000006  881  0X4A14
  3 @9.67.107.7         7.7.7.7         0X80000003  880  0X4618
  3 @9.67.108.2         4.4.4.4         0X80000003 1667  0XBDB1
  3 @9.67.108.4         4.4.4.4         0X80000003 1658  0XB3B8
  4 @2.2.2.2            4.4.4.4         0X80000003 1658  0XAC74
  4 @5.5.5.5            7.7.7.7         0X80000003  880  0XC741
               # ADVERTISEMENTS:       14
               CHECKSUM TOTAL:         0X884B0
```

Displaying current, run-time statistics and parameters for OSPF interfaces

To display current, run-time statistics and parameters for OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,INTERFACE
EZZ7849I INTERFACES 763
```

```
IFC ADDRESS     PHYS         ASSOC. AREA     TYPE   STATE  #NBRS  #ADJS
7.7.7.7         VIPA1A       1.1.1.1         VIPA    N/A    N/A    N/A
9.67.104.7      NBMA7        1.1.1.1         MULTI     1      3      0
9.67.100.7      CTC7TO8      1.1.1.1         P-P      16      1      1
9.67.102.7      CTC7TO3      1.1.1.1         P-P      16      1      1
9.67.106.7      CTC7TO4      1.1.1.1         P-P      16      1      1
9.67.107.7      CTC7TO5      0.0.0.0         P-P      16      1      1
UNNUMBERED      VL/0         0.0.0.0         VLINK    16      1      1
```

Displaying current, run-time statistics and parameters for a specific OSPF interface

To display current, run-time statistics and parameters for a specific OSPF interface, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,IF,NAME=CTC7TO4
EZZ7850I INTERFACE DETAILS 769
                INTERFACE ADDRESS:      9.67.106.7
                ATTACHED AREA:          1.1.1.1
                PHYSICAL INTERFACE:     CTC7TO4
                INTERFACE MASK:         255.255.255.0
                INTERFACE TYPE:         P-P
                STATE:                  16
                DESIGNATED ROUTER:      N/A
                BACKUP DR:              N/A

  DR PRIORITY:     N/A  HELLO INTERVAL:   10  RXMT INTERVAL:     5
  DEAD INTERVAL:    40  TX DELAY:          1  POLL INTERVAL:     0
  DEMAND CIRCUIT:  OFF  HELLO SUPPRESS:  OFF  SUPPRESS REQ:    OFF
  MAX PKT SIZE:   1024  TOS 0 COST:        1  DB_EX INTERVAL:   40
  AUTH TYPE:  PASSWORD

  # NEIGHBORS:       1  # ADJACENCIES:     1  # FULL ADJS.:      1
  # MCAST FLOODS:   15  # MCAST ACKS:      4

  NETWORK CAPABILITIES:
   POINT-TO-POINT
   DEMAND-CIRCUITS
```

Displaying current, run-time statistics and parameters for OSPF neighbors

To display current, run-time statistics and parameters for OSPF neighbors, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,NBR
EZZ7851I NEIGHBOR SUMMARY 771
NEIGHBOR ADDR   NEIGHBOR ID     STATE  LSRXL DBSUM LSREQ HSUP IFC
9.67.104.16     0.0.0.0             1     0     0     0  OFF NBMA7
9.67.104.25     0.0.0.0             1     0     0     0  OFF NBMA7
9.67.104.15     0.0.0.0             1     0     0     0  OFF NBMA7
9.67.100.8      8.8.8.8           128     0     0     0  OFF CTC7TO8
9.67.102.3      3.3.3.3           128     0     0     0  OFF CTC7TO3
9.67.106.4      4.4.4.4           128     0     0     0  OFF CTC7TO4
9.67.107.5      5.5.5.5           128     0     0     0  OFF CTC7TO5
VL/0            4.4.4.4           128     0     0     0  OFF *
```

Displaying current run-time statistics and parameters for a specific OSPF neighbor

To display current run-time statistics and parameters for a specific OSPF neighbor, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,NBR,IPADDR=9.67.106.4
EZZ7852I NEIGHBOR DETAILS 779
                NEIGHBOR IP ADDRESS:    9.67.106.4
                OSPF ROUTER ID:         4.4.4.4
                NEIGHBOR STATE:         128
                PHYSICAL INTERFACE:     CTC7TO4
                DR CHOICE:              0.0.0.0
                BACKUP CHOICE:          0.0.0.0
                DR PRIORITY:            1
                NBR OPTIONS:            E
```

```
  DB SUMM QLEN:      0  LS RXMT QLEN:      0  LS REQ QLEN:       0
  LAST HELLO:        4  NO HELLO:        OFF
  # LS RXMITS:       1  # DIRECT ACKS:     0  # DUP LS RCVD:     6
  # OLD LS RCVD:     0  # DUP ACKS RCVD:   1  # NBR LOSSES:      0
  # ADJ. RESETS:     0
```

Displaying routes to other routers that have been calculated by OSPF

To display routes to other routers that have been calculated by OSPF, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,ROUTERS
EZZ7855I OSPF ROUTERS 781
DTYPE RTYPE DESTINATION       AREA           COST       NEXT HOP(S)
ASBR  SPF   2.2.2.2          0.0.0.0         2         9.67.106.4
  BR  SPF   4.4.4.4          0.0.0.0         1         9.67.106.4
ASBR  SPF   5.5.5.5          0.0.0.0         1         9.67.107.5
ASBR  SPF   3.3.3.3          1.1.1.1         1         9.67.102.3
  BR  SPF   4.4.4.4          1.1.1.1         1         9.67.106.4
ASBR  SPF   8.8.8.8          1.1.1.1         1         9.67.100.8
```

Displaying the number of LSAs currently in the link state database

To display the number of LSAs currently in the link state database, categorized by type, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,DBSIZE
EZZ7854I LINK STATE DATABASE SIZE 783
                # ROUTER-LSAS:            8
                # NETWORK-LSAS:           0
                # SUMMARY-LSAS:           37
                # SUMMARY ROUTER-LSAS:    7
                # AS EXTERNAL-LSAS:       18
                # INTRA-AREA ROUTES:      24
                # INTER-AREA ROUTES:      1
                # TYPE 1 EXTERNAL ROUTES: 0
                # TYPE 2 EXTERNAL ROUTES: 0
```

Displaying statistics generated by the OSPF routing protocol

To display statistics generated by the OSPF routing protocol, enter the following command:

```
D TCPIP,TCPCS7,OMP,OSPF,STATS
EZZ7856I OSPF STATISTICS 785
                OSPF ROUTER ID:         7.7.7.7 (*OSPF)
                EXTERNAL COMPARISON:    TYPE 2
                AS BOUNDARY CAPABILITY: YES
                IMPORT EXTERNAL ROUTES: RIP SUB
                ORIG. DEFAULT ROUTE:    YES
                DEFAULT ROUTE COST:     (1, TYPE 2)
                DEFAULT FORWARD. ADDR.: 0.0.0.0
                LEARN HIGHER COST DFLT: NO
ATTACHED AREAS:                  2  OSPF PACKETS RCVD:              821
OSPF PACKETS RCVD W/ERRS:        0  TRANSIT NODES ALLOCATED:         55
TRANSIT NODES FREED:            47  LS ADV. ALLOCATED:              263
LS ADV. FREED:                 201  QUEUE HEADERS ALLOC:             96
QUEUE HEADERS AVAIL:            96  MAXIMUM LSA SIZE:               976
# DIJKSTRA RUNS:                9  INCREMENTAL SUMM. UPDATES:        4
INCREMENTAL VL UPDATES:          0  MULTICAST PKTS SENT:            746
UNICAST PKTS SENT:             107  LS ADV. AGED OUT:                 0
LS ADV. FLUSHED:                22  PTRS TO INVALID LS ADV:           0
INCREMENTAL EXT. UPDATES:       49
```

Displaying all of the RIP configuration information

To display all of the RIP configuration information, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,LIST,ALL
EZZ7843I RIP CONFIGURATION 800
TRACE: 0, DEBUG: 0, SADEBUG LEVEL: 0
STACK AFFINITY: TCPCS7
RIP: ENABLED
```

```
RIP DEFAULT ORIGINATION: ALWAYS, COST = 1
PER-INTERFACE ADDRESS FLAGS:
CTC7TO6          9.67.103.7      RIP-2 MULTICAST.
                                 SEND NET AND SUBNET ROUTES
                                 RECEIVE NO DYNAMIC HOST ROUTES
                                 RIP INTERFACE INPUT METRIC: 1
                                 RIP INTERFACE OUTPUT METRIC: 0

EZZ7844I RIP ROUTE ACCEPTANCE
ACCEPT RIP UPDATES ALWAYS FOR:
  30.1.1.8          30.1.1.4
IGNORE RIP UPDATES FROM:
   NONE
```

Displaying information about configured RIP interfaces

To display information about configured RIP interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,LIST,IFS
EZZ7843I RIP CONFIGURATION 806
TRACE: 0, DEBUG: 0, SADEBUG LEVEL: 0
STACK AFFINITY: TCPCS7
RIP: ENABLED
RIP DEFAULT ORIGINATION: ALWAYS, COST = 1
PER-INTERFACE ADDRESS FLAGS:
CTC7TO6          9.67.103.7      RIP-2 MULTICAST.
                                 SEND NET AND SUBNET ROUTES
                                 RECEIVE NO DYNAMIC HOST ROUTES
                                 RIP INTERFACE INPUT METRIC: 1
                                 RIP INTERFACE OUTPUT METRIC: 0
                                 RIP RECEIVE CONTROL: ANY
```

Displaying the routes to be unconditionally accepted

To display the routes to be unconditionally accepted, as configured with the Accept_RIP_Route statement, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,LIST,ACCEPTED
EZZ7844I RIP ROUTE ACCEPTANCE 808
ACCEPT RIP UPDATES ALWAYS FOR:
  30.1.1.8          30.1.1.4
```

Displaying current run-time information about RIP interfaces

To display current, run-time information about RIP interfaces, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,IF
EZZ7859I RIP INTERFACES 810
IFC ADDRESS      IFC NAME         SUBNET MASK     MTU   DESTINATION
9.67.103.7       CTC7TO6          255.255.255.0   1024  0.0.0.0
```

Displaying current run-time information about a specific RIP interface

To display current, run-time information about a specific RIP interface, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,IF,NAME=CTC7TO6
EZZ7860I RIP INTERFACE DETAILS 812
INTERFACE ADDRESS:      9.67.103.7
INTERFACE NAME:         CTC7TO6
SUBNET MASK:            255.255.255.0
MTU                     1024
DESTINATION ADDRESS:    0.0.0.0

RIP VERSION:            2     SEND POIS. REV. ROUTES: YES
IN METRIC:              1     OUT METRIC:             0
RECEIVE NET ROUTES:     YES   RECEIVE SUBNET ROUTES:  YES
RECEIVE HOST ROUTES:    NO    SEND DEFAULT ROUTES:    NO
SEND NET ROUTES:        YES   SEND SUBNET ROUTES:     YES
SEND STATIC ROUTES:     NO    SEND HOST ROUTES:       NO

SEND ONLY: ALL
```

```
RIP RECEIVE CONTROL:    ANY
```

Displaying the global RIP filters

To display the global RIP filters, enter the following command:

```
D TCPIP,TCPCS7,OMP,RIP,FILTERS
EZZ8016I GLOBAL RIP FILTERS 684
SEND ONLY: ALL

IGNORE RIP UPDATES FROM:
                     9.67.103.10              9.67.103.9

FILTERS: NOSEND          10.1.1.0           255.255.255.0
```

Displaying the routes in the OMPROUTE main routing table

To display all of the routes in the OMPROUTE main routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE
EZZ7847I ROUTING TABLE 796
TYPE   DEST NET         MASK      COST    AGE     NEXT HOP(S)

SBNT   2.0.0.0          FF000000  1       1368    NONE
 SPF   2.2.2.0          FFFFFFFC  3       1380    9.67.106.4
 SPF   2.2.2.2          FFFFFFFF  3       1380    9.67.106.4
SBNT   3.0.0.0          FF000000  1       1549    NONE
 SPF   3.3.3.0          FFFFFFFC  2       1561    9.67.102.3
 SPF   3.3.3.3          FFFFFFFF  2       1561    9.67.102.3
SBNT   4.0.0.0          FF000000  1       1549    NONE
 SPF   4.4.4.4          FFFFFFFC  2       1561    9.67.106.4
 SPF   4.4.4.4          FFFFFFFF  2       1561    9.67.106.4
SBNT   5.0.0.0          FF000000  1       1549    NONE
 SPF   5.5.5.4          FFFFFFFC  2       1567    9.67.107.5
 SPF   5.5.5.5          FFFFFFFF  2       1567    9.67.107.5
SBNT   6.0.0.0          FF000000  1       1549    NONE
 RIP   6.6.6.4          FFFFFFFC  2       30      9.67.103.6
SBNT   7.0.0.0          FF000000  1       1368    NONE
SPIA*  7.7.7.4          FFFFFFFC  3       1380    9.67.106.4
 DIR*  7.7.7.7          FFFFFFFF  1       1574    VIPA1A
SBNT   8.0.0.0          FF000000  1       1549    NONE
 SPF   8.8.8.8          FFFFFFFC  2       1545    9.67.100.8
 SPF   8.8.8.8          FFFFFFFF  2       1545    9.67.100.8
SBNT   9.0.0.0          FF000000  1       1368    NONE
 DIR*  9.67.100.0       FFFFFF00  1       1576    9.67.100.7
 SPF   9.67.100.7       FFFFFFFF  2       1545    CTC7TO8
 SPF   9.67.100.8       FFFFFFFF  1       1572    9.67.100.8
 SPF   9.67.101.3       FFFFFFFF  2       1561    9.67.106.4
 SPF   9.67.101.4       FFFFFFFF  2       1561    9.67.102.3
 DIR*  9.67.102.0       FFFFFF00  1       1575    9.67.102.7
 SPF   9.67.102.3       FFFFFFFF  1       1566    9.67.102.3
 SPF   9.67.102.7       FFFFFFFF  2       1561    CTC7TO3
 DIR*  9.67.103.0       FFFFFF00  1       1575    9.67.103.7
 RIP   9.67.103.6       FFFFFFFF  1       30      9.67.103.6
 SPF   9.67.105.4       FFFFFFFF  2       1545    9.67.100.8
 SPF   9.67.105.8       FFFFFFFF  2       1561    9.67.106.4
 DIR*  9.67.106.0       FFFFFF00  1       1576    9.67.106.7
 SPF   9.67.106.4       FFFFFFFF  1       1566    9.67.106.4
 SPF   9.67.106.7       FFFFFFFF  2       1561    CTC7TO4
 DIR*  9.67.107.0       FFFFFF00  1       1577    9.67.107.7
 SPF   9.67.107.5       FFFFFFFF  1       1574    9.67.107.5
 SPF   9.67.107.7       FFFFFFFF  2       1566    CTC7TO5
 SPF   9.67.108.2       FFFFFFFF  2       1380    9.67.106.4
 SPF   9.67.108.4       FFFFFFFF  3       1380    9.67.106.4
SBNT   10.0.0.0         FF000000  1       1368    NONE
SPE2   10.1.1.0         FFFFFF00  0       1379    9.67.106.4
SPE2   10.1.1.1         FFFFFFFF  0       1379    9.67.106.4
SBNT   20.0.0.0         FF000000  1       1549    NONE
SPE2   20.1.1.0         FFFFFF00  0       1379    9.67.107.5
SPE2   20.1.1.1         FFFFFFFF  0       1379    9.67.107.5
 RIP   30.0.0.0         FF000000  2       30      9.67.103.6
 RIP   30.1.1.0         FFFFFF00  2       30      9.67.103.6
 RIP % 30.1.1.4         FFFFFFFF  2       30      9.67.103.6
 RIP % 30.1.1.8         FFFFFFFF  2       30      9.67.103.6
SPE2   130.200.0.0      FFFF0000  0       1379    9.67.100.8      (2)
SPE2   130.200.1.1      FFFFFFFF  0       1379    9.67.102.3
```

```
SPE2   130.200.1.18     FFFFFFFF  0       1379    9.67.100.8
SPE2   130.201.0.0      FFFF0000  0       1379    9.67.100.8       (2)
SPE2   130.202.0.0      FFFF0000  0       1379    9.67.100.8       (2)
                      0 NETS DELETED, 4 NETS INACTIVE
```

Displaying the routes to a specific destination in the main routing table

To display information about the routes to a specific destination that are in the main routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE,DEST=130.201.0.0
EZZ7874I ROUTE EXPANSION 798
DESTINATION:    130.201.0.0
MASK:           255.255.0.0
ROUTE TYPE:     SPE2
DISTANCE:       0
AGE:            1485
NEXT HOP(S):    9.67.100.8        (CTC7TO8)
                9.67.102.3        (CTC7TO3)
```

Displaying the routes in all OMPROUTE IPv4 policy-based routing tables

To display all of the routes in all OMPROUTE policy-based routing tables, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE,PR=ALL
EZZ7847I ROUTING TABLE 154
TABLE NAME: SECHIGH
TYPE   DEST NET         MASK      COST    AGE     NEXT HOP(S)

SBNT   2.0.0.0          FF000000  1       192     NONE
SPIA   2.2.2.0          FFFFFFFC  4       201     9.67.100.8
SPIA   2.2.2.2          FFFFFFFF  4       201     9.67.100.8
SBNT   3.0.0.0          FF000000  1       192     NONE
 SPF   3.3.3.0          FFFFFFF0  103     201     9.67.100.8
 SPF   3.3.3.3          FFFFFFFF  103     201     9.67.100.8
SBNT   4.0.0.0          FF000000  1       192     NONE
 SPF   4.4.4.4          FFFFFFFC  3       201     9.67.100.8
 SPF   4.4.4.4          FFFFFFFF  3       201     9.67.100.8
SBNT   7.0.0.0          FF000000  1       192     NONE
STAT*  7.7.77.77        FFFFFFFF  0       203     8.8.88.8
SBNT   8.0.0.0          FF000000  1       192     NONE
 SPF   8.8.8.8          FFFFFFFC  2       201     9.67.100.8
 SPF   8.8.8.8          FFFFFFFF  2       201     9.67.100.8
SBNT   9.0.0.0          FF000000  1       192     NONE
 DIR*  9.67.100.0       FFFFFF00  1       201     9.67.100.7
 SPF   9.67.100.8       FFFFFFFF  1       201     9.67.100.8
 SPF   9.67.101.3       FFFFFFFF  102     201     9.67.100.8
 SPF   9.67.101.4       FFFFFFFF  103     201     9.67.100.8
 SPF   9.67.105.4       FFFFFFFF  2       201     9.67.100.8
 SPF   9.67.105.8       FFFFFFFF  7       201     9.67.100.8
SPIA   9.67.106.4       FFFFFFFF  4       201     9.67.100.8
SPIA   9.67.107.7       FFFFFFFF  5       201     9.67.100.8
SPIA   9.67.108.2       FFFFFFFF  3       201     9.67.100.8
SBNT   10.0.0.0         FF000000  1       192     NONE
SPE2   10.1.1.0         FFFFFF00  1       201     9.67.100.8
SPE2   10.1.1.1         FFFFFFFF  1       201     9.67.100.8
SPE2   130.200.0.0      FFFF0000  0       192     9.67.100.8
SPE2   130.200.1.1      FFFFFFFF  0       201     9.67.100.8
SPE2   130.200.1.18     FFFFFFFF  0       201     9.67.100.8
SPE2   130.201.0.0      FFFF0000  0       201     9.67.100.8
SPE2   130.202.0.0      FFFF0000  0       201     9.67.100.8
                      0 NETS DELETED
DYNAMIC ROUTING PARAMETERS
  INTERFACE:  CTC7TO8          NEXT HOP: ANY

TABLE NAME: SECLOW
TYPE   DEST NET         MASK      COST    AGE     NEXT HOP(S)

SBNT   2.0.0.0          FF000000  1       192     NONE
SPIA   2.2.2.0          FFFFFFFC  4       201     9.67.102.3
SPIA   2.2.2.2          FFFFFFFF  4       201     9.67.102.3
SBNT   3.0.0.0          FF000000  1       192     NONE
 SPF   3.3.3.0          FFFFFFF0  2       201     9.67.102.3
 SPF   3.3.3.3          FFFFFFFF  2       201     9.67.102.3
SBNT   4.0.0.0          FF000000  1       192     NONE
 SPF   4.4.4.4          FFFFFFFC  3       201     9.67.102.3
```

```
 SPF   4.4.4.4          FFFFFFFF  3       201     9.67.102.3
SBNT   7.0.0.0          FF000000  1       192     NONE
STAT*  7.7.7.77         FFFFFFFF  0       203     3.3.33.3
SBNT   8.0.0.0          FF000000  1       192     NONE
 SPF   8.8.8.8          FFFFFFFC  8       201     9.67.102.3
 SPF   8.8.8.8          FFFFFFFF  8       201     9.67.102.3
SBNT   9.0.0.0          FF000000  1       192     NONE
 SPF   9.67.101.3       FFFFFFFF  102     201     9.67.102.3
 SPF   9.67.101.4       FFFFFFFF  2       201     9.67.102.3
 DIR*  9.67.102.0       FFFFFF00  1       201     9.67.102.7
 SPF   9.67.102.3       FFFFFFFF  1       201     9.67.102.3
 SPF   9.67.105.4       FFFFFFFF  8       201     9.67.102.3
 SPF   9.67.105.8       FFFFFFFF  7       201     9.67.102.3
SPIA   9.67.106.4       FFFFFFFF  4       201     9.67.102.3
SPIA   9.67.107.7       FFFFFFFF  5       201     9.67.102.3
SPIA   9.67.108.2       FFFFFFFF  3       201     9.67.102.3
SBNT   10.0.0.0         FF000000  1       192     NONE
SPE2   10.1.1.0         FFFFFF00  1       201     9.67.102.3
SPE2   10.1.1.1         FFFFFFFF  1       201     9.67.102.3
SPE2   130.200.0.0      FFFF0000  0       192     9.67.102.3
SPE2   130.200.1.1      FFFFFFFF  0       201     9.67.102.3
SPE2   130.200.1.18     FFFFFFFF  0       201     9.67.102.3
SPE2   130.201.0.0      FFFF0000  0       201     9.67.102.3
SPE2   130.202.0.0      FFFF0000  0       201     9.67.102.3
                       0 NETS DELETED
DYNAMIC ROUTING PARAMETERS
  INTERFACE:  CTC7TO3          NEXT HOP: ANY
```

Result: If a policy-based route table is configured with no IPv4 dynamic routing parameters, OMPROUTE has no knowledge of that route table for IPv4. The route table is not included in this display. If no policy-based route tables are configured with IPv4 dynamic routing parameters, message EZZ8150I is issued.

Displaying the routes in an OMPROUTE IPv4 policy-based routing table

To display all of the routes in an OMPROUTE policy-based routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE,PR=SECHIGH
EZZ7847I ROUTING TABLE 154
TABLE NAME: SECHIGH
TYPE   DEST NET         MASK      COST    AGE     NEXT HOP(S)

SBNT   2.0.0.0          FF000000  1       192     NONE
SPIA   2.2.2.0          FFFFFFFC  4       201     9.67.100.8
SPIA   2.2.2.2          FFFFFFFF  4       201     9.67.100.8
SBNT   3.0.0.0          FF000000  1       192     NONE
 SPF   3.3.3.0          FFFFFFF0  103     201     9.67.100.8
 SPF   3.3.3.3          FFFFFFFF  103     201     9.67.100.8
SBNT   4.0.0.0          FF000000  1       192     NONE
 SPF   4.4.4.4          FFFFFFFC  3       201     9.67.100.8
 SPF   4.4.4.4          FFFFFFFF  3       201     9.67.100.8
SBNT   7.0.0.0          FF000000  1       192     NONE
STAT*  7.7.77.77        FFFFFFFF  0       203     8.8.88.8
SBNT   8.0.0.0          FF000000  1       192     NONE
 SPF   8.8.8.8          FFFFFFFC  2       201     9.67.100.8
 SPF   8.8.8.8          FFFFFFFF  2       201     9.67.100.8
SBNT   9.0.0.0          FF000000  1       192     NONE
 DIR*  9.67.100.0       FFFFFF00  1       201     9.67.100.7
 SPF   9.67.100.8       FFFFFFFF  1       201     9.67.100.8
 SPF   9.67.101.3       FFFFFFFF  102     201     9.67.100.8
 SPF   9.67.101.4       FFFFFFFF  103     201     9.67.100.8
 SPF   9.67.105.4       FFFFFFFF  2       201     9.67.100.8
 SPF   9.67.105.8       FFFFFFFF  7       201     9.67.100.8
SPIA   9.67.106.4       FFFFFFFF  4       201     9.67.100.8
SPIA   9.67.107.7       FFFFFFFF  5       201     9.67.100.8
SPIA   9.67.108.2       FFFFFFFF  3       201     9.67.100.8
SBNT   10.0.0.0         FF000000  1       192     NONE
SPE2   10.1.1.0         FFFFFF00  1       201     9.67.100.8
SPE2   10.1.1.1         FFFFFFFF  1       201     9.67.100.8
SPE2   130.200.0.0      FFFF0000  0       192     9.67.100.8
SPE2   130.200.1.1      FFFFFFFF  0       201     9.67.100.8
SPE2   130.200.1.18     FFFFFFFF  0       201     9.67.100.8
SPE2   130.201.0.0      FFFF0000  0       201     9.67.100.8
SPE2   130.202.0.0      FFFF0000  0       201     9.67.100.8
                       0 NETS DELETED
DYNAMIC ROUTING PARAMETERS
  INTERFACE:  CTC7TO8          NEXT HOP: ANY
```

Result: If the policy-based route table is configured with no IPv4 dynamic routing parameters, OMPROUTE has no knowledge of the route table for IPv4. If no policy-based route tables are configured with IPv4 dynamic routing parameters, message EZZ8150I is issued.

Displaying the routes to a specific destination in an IPv4 policy-based routing table

To display information about the routes to a specific destination that are in a policy-based routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RTTABLE,PR=SECHIGH,DEST=130.201.0.0
EZZ7874I ROUTE EXPANSION 165
TABLE NAME: SECHIGH
DESTINATION:    130.201.0.0
MASK:           255.255.0.0
ROUTE TYPE:     SPE2
DISTANCE:       0
AGE:            548
NEXT HOP(S):    9.67.100.8        (CTC7TO8)
```

Result: If the policy-based route table is configured with no IPv4 dynamic routing parameters, OMPROUTE has no knowledge of the route table for IPv4. If no policy-based route tables are configured with IPv4 dynamic routing parameters, message EZZ8150I is issued.

Displaying all of the generic configuration information

To display all of the IPv4 configuration information that is not related to any routing protocol, enter the following command:

```
D TCPIP,TCPCS3,OMP,GENERIC,LIST,ALL
EZZ8053I IPV4 GENERIC CONFIGURATION
TRACE: 2, DEBUG: 3, SADEBUG LEVEL: 0
IPV4 TRACE DESTINATION: /TMP/AMPROUT3.DBG
STACK AFFINITY: TCPCS3

EZZ8056I IPV4 GEN INT CONFIGURATION
IFC NAME         IFC ADDRESS     SUBNET MASK       MTU DESTADDR
NSQDIO3L         9.67.120.3      255.255.255.0     576  N/A
CTC3TO4          9.67.101.3      255.255.255.0   10000  9.67.101.4
```

Displaying information about configured generic interfaces

To display information about configured generic interfaces (that is, interfaces defined to OMPROUTE with the INTERFACE statement, or not defined to OMPROUTE but learned from the stack), enter the following command:

```
D TCPIP,TCPCS3,OMP,GENERIC,LIST,IFS
EZZ8056I IPV4 GEN INT CONFIGURATION
IFC NAME         IFC ADDRESS     SUBNET MASK       MTU DESTADDR
NSQDIO3L         9.67.120.3      255.255.255.0     576  N/A
CTC3TO4          9.67.101.3      255.255.255.0   10000  9.67.101.4
```

Displaying current run-time information about generic interfaces

To display current run-time information about generic interfaces, enter the following command:

```
D TCPIP,TCPCS3,OMP,GENERIC,IFS
EZZ8060I IPV4 GENERIC INTERFACES
IFC NAME         IFC ADDRESS     SUBNET MASK       MTU  CFG  IGN
NSQDIO3L         9.67.120.3      255.255.255.0     576  YES   NO
CTC3TO1          130.200.1.3     N/A               N/A   NO  YES
VIPA03           3.3.3.103       N/A               N/A   NO  YES
CTC3TO4          9.67.101.3      255.255.255.0   10000  YES   NO
```

Verification of OMPROUTE IPv6 configuration and state

The following topics show sample output from each of the commands that can be used to display OMPROUTE IPv6 information. The syntax of these DISPLAY commands, as well as detailed information about the data displayed, can be found in z/OS Communications Server: IP System Administrator's Commands.

Displaying all IPv6 OSPF information

To display a comprehensive list of IPv6 OSPF information, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,ALL
EZZ7970I IPV6 OSPF INFORMATION 322
TRACE6: 0, DEBUG6: 0
STACK AFFINITY            TCPCS67
IPV6 OSPF PROTOCOL:       ENABLED
IPV6 OSPF ROUTER ID:      67.67.67.67 (*IPV6_OSPF)
DFLT IPV6 OSPF INST ID:   0
EXTERNAL COMPARISON:      TYPE 2
AS BOUNDARY CAPABILITY:   ENABLED
IMPORT EXTERNAL ROUTES:   RIP
ORIG. DEFAULT ROUTE:      NO
DEMAND CIRCUITS:          ENABLED

EZZ7973I IPV6 OSPF AREAS
AREA ID         STUB DFLT-COST IMPORT-PREF DEMAND IFCS NETS RTRS ABRS
6.6.6.6          NO        N/A     N/A       OFF     2    1    4    2
0.0.0.0          NO        N/A     N/A       OFF     2    0    4    2

--AREA RANGES--
AREA ID         ADVERTISE  PREFIX
6.6.6.6             NO     2001:DB8:0:101::/64

EZZ7958I IPV6 OSPF INTERFACES
NAME             AREA            TYPE   STATE COST HELLO DEAD NBRS ADJS
VIPA1A6          6.6.6.6         VIPA    N/A     1  N/A   N/A  N/A  N/A
MPCPTP7TO5       0.0.0.0         P-2-MP   16     1   10    40    1    1
NSQDIO1L6        6.6.6.6         BRDCST   32     1   10    40    3    2
VL/0             0.0.0.0         VLINK    16     1   30   180    1    1

EZZ7972I IPV6 OSPF VIRTUAL LINKS
ENDPOINT        TRANSIT AREA   STATE COST HELLO DEAD NBRS ADJS
64.64.64.64     6.6.6.6          16     1   30   180    1    1

EZZ8129I IPV6 OSPF NEIGHBORS
ROUTER ID       STATE LSRXL DBSUM LSREQ HSUP RTR-PRI IFC
65.65.65.65       128     0     0     0  OFF       1 MPCPTP7TO5
64.64.64.64       128     0     0     0  OFF       1 NSQDIO1L6
63.63.63.63       128     0     0     0  OFF       1 NSQDIO1L6
68.68.68.68       128     0     0     0  OFF       1 NSQDIO1L6
64.64.64.64       128     0     0     0  OFF       1 *
```

Displaying IPv6 OSPF area statistics and parameters

To display the statistics and parameters for all IPv6 OSPF areas attached to the router, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,AREASUM
EZZ7973I IPV6 OSPF AREAS 536
AREA ID         STUB DFLT-COST IMPORT-PREF DEMAND IFCS NETS RTRS ABRS
6.6.6.6          NO        N/A     N/A       OFF     2    1    4    2
0.0.0.0          NO        N/A     N/A       OFF     2    0    4    2

--AREA RANGES--
AREA ID         ADVERTISE  PREFIX
6.6.6.6             NO     2001:DB8:0:101::/64
```

Displaying IPv6 OSPF interface statistics and parameters

To display current, run-time statistics and parameters related to IPv6 OSPF interfaces, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,IFS
EZZ7958I IPV6 OSPF INTERFACES 575
NAME             AREA            TYPE   STATE COST HELLO DEAD NBRS ADJS
VIPA1A6          6.6.6.6         VIPA     N/A    1  N/A   N/A  N/A  N/A
MPCPTP7TO5       0.0.0.0         P-2-MP    16    1   10    40    1    1
NSQDIO1L6        6.6.6.6         BRDCST    32    1   10    40    3    2
VL/0             0.0.0.0         VLINK     16    1   30   180    1    1
```

Displaying statistics and parameters for a specific IPv6 OSPF interface

To display current, run-time statistics and parameters for a specific IPv6 OSPF interface, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,IF,NAME=NSQDIO1L6
EZZ7959I IPV6 OSPF INTERFACE DETAIL 677
INTERFACE NAME:     NSQDIO1L6
INTERFACE ID:       20
INSTANCE ID:        0
INTERFACE ADDRESS:  FE80::7
                    2001:DB8:0:120::7
INTERFACE PREFIX:   STAT 2001:DB8:0:120::/64
ATTACHED AREA:      6.6.6.6
INTERFACE TYPE:     BRDCST
STATE:              32
DESIGNATED ROUTER:  68.68.68.68
BACKUP DR:          64.64.64.64

DR PRIORITY:       1  HELLO INTERVAL:   10  RXMT INTERVAL:     5
DEAD INTERVAL:    40  TX DELAY:          1  POLL INTERVAL:   N/A
DEMAND CIRCUIT:  OFF  HELLO SUPPRESS:  N/A  SUPPRESS REQ:    N/A
MTU:            9000  COST:             1  DB_EX INTERVAL:   40

# NEIGHBORS:       3  # ADJACENCIES:     2  # FULL ADJS.:      2
# MCAST FLOODS:    7  # MCAST ACKS:      9

NETWORK CAPABILITIES:
 BROADCAST
 DEMAND-CIRCUITS
 MULTICAST
```

Displaying IPv6 OSPF virtual link statistics and parameters

To display current, run-time statistics and parameters related to IPv6 OSPF virtual links, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,VLINK
EZZ7972I IPV6 OSPF VIRTUAL LINKS 703
ENDPOINT        TRANSIT AREA   STATE COST HELLO DEAD NBRS ADJS
64.64.64.64     6.6.6.6          16     1   30   180    1    1
```

Displaying statistics and parameters for a specific IPv6 OSPF virtual link

To display current, run-time statistics and parameters for a specific IPv6 OSPF virtual link, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,VLINK,ENDPT=64.64.64.64
EZZ7971I IPV6 VIRTUAL LINK DETAILS 713
VIRTUAL LINK ENDPOINT:     64.64.64.64
PHYSICAL INTERFACE NAME:   NSQDIO1L6
VL TRANSIT AREA:           6.6.6.6
STATE:                     16

HELLO INTERVAL:     30  DEAD INTERVAL:     180  DB_EX INTERVAL:    180
RXMT INTERVAL:      10  TX DELAY:            5  COST:                1
DEMAND CIRCUIT:     ON  HELLO SUPPRESS:    OFF  SUPPRESS REQ:       ON
```

```
# NEIGHBORS:         1  # ADJACENCIES:       1  # FULL ADJS.:        1
```

Displaying IPv6 OSPF neighbor statistics and parameters

To display the statistics and parameters related to IPv6 OSPF neighbors, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,NBRS
EZZ8129I IPV6 OSPF NEIGHBORS 715
ROUTER ID       STATE LSRXL DBSUM LSREQ HSUP RTR-PRI IFC
65.65.65.65       128     0     0     0  OFF       1 MPCPTP7TO5
63.63.63.63         8     0     0     0  OFF       1 NSQDIO1L6
64.64.64.64       128     0     0     0  OFF       1 NSQDIO1L6
68.68.68.68       128     0     0     0  OFF       1 NSQDIO1L6
64.64.64.64       128     0     0     0  OFF       1 *
```

Tip: On multiaccess media (LANs), attached routers become adjacent only with the designated router and the backup designated router. Routers that are not performing a designated router role do not become adjacent to each other on LAN networks. Therefore, in this example, the state of 8 for 63.63.63.63 does not necessarily represent a problem or an incomplete adjacency. You can conclude that 64.64.64.64 and 68.68.68.68 are the designated routers, and 63.63.63.63 is simply another router on the network. State 8 is the final state in this case.

Displaying statistics and parameters for a specific IPv6 OSPF neighbor

To display current, run-time statistics and parameters for a specific IPv6 OSPF neighbor, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,NBR,ID=64.64.64.64,IFNAME=NSQDIO1L6
EZZ8130I IPV6 OSPF NEIGHBOR DETAILS 737
NEIGHBOR IP ADDRESS:    FE80::4
OSPF ROUTER ID:         64.64.64.64
NEIGHBOR STATE:         128
PHYSICAL INTERFACE:     NSQDIO1L6
DR CHOICE:              68.68.68.68
BACKUP CHOICE:          64.64.64.64
DR PRIORITY:            1
NBR OPTIONS:            (0X00)

DB SUMM QLEN:      0  LS RXMT QLEN:      0  LS REQ QLEN:       0
LAST HELLO:        5  NO HELLO:        OFF
# LS RXMITS:       1  # DIRECT ACKS:     5  # DUP LS RCVD:     4
# OLD LS RCVD:     0  # DUP ACKS RCVD:   3  # ADJ. RESETS:     1
```

Displaying IPv6 OSPF link state database statistics

To display the number of LSAs currently in the link state database, categorized by type, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,DBSIZE
EZZ8128I IPV6 OSPF LS DATABASE SIZE 841
# ROUTER-LSAS:              8
# NETWORK-LSAS:             1
# INTER-AREA PREFIX LSAS:   50
# INTER-AREA ROUTER LSAS:   6
# AS EXTERNAL-LSAS:         6
# LINK LSAS:                6
# INTRA-AREA PREFIX LSAS:   21
# UNKNOWN LSAS:             0
# INTRA-AREA ROUTES:        24
# INTER-AREA ROUTES:        0
# TYPE 1 EXTERNAL ROUTES:   0
# TYPE 2 EXTERNAL ROUTES:   0
```

Displaying IPv6 OSPF link state advertisement

To display the contents of a single link state advertisement contained in the IPv6 OSPF database, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,LSA,LSTYPE=2001,LSID=0,ORIG=64.64.64.64,
AREAID=6.6.6.6
EZZ7880I LSA DETAILS 834
        LS AGE:          61
        LS TYPE:         0X2001 (ROUTER LSA)
        LS ID:           0
        LS ORIGINATOR:   64.64.64.64
        LS SEQUENCE NO:  0X8000000F
        LS CHECKSUM:     0X3886
        LS LENGTH:       40
        ROUTER TYPE:     (0X01) ABR
        LS OPTIONS:      (0X000033) V6,E,R,DC
INTERFACES:
 TYPE  METRIC  INTERFACE ID   NBR INTERFACE ID   NBR ROUTER ID
    2       1            16                 14     68.68.68.68
```

Displaying IPv6 OSPF external advertisements

To display the AS external advertisements belonging to the IPv6 OSPF routing domain, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,EXTERNAL
EZZ8127I IPV6 OSPF AS EXTERNAL LSDB 555
              AS EXTERNAL LSAS (LS TYPE=4005)
LS ORIGINATOR   LS ID      SEQNO       AGE PREFIX
67.67.67.67     5          0X80000001  565 6:6:6:6:6:6:6:6/128
67.67.67.67     6          0X80000001  561 2001:DB8:0:A1C::6/128
67.67.67.67     7          0X80000001  558 2001:DB8:0:103::6/128
67.67.67.67     8          0X80000001  222 2001:DB8:0:A10::/60
67.67.67.67     9          0X80000001  222 2001:DB8:0:A1B::/64
67.67.67.67     10         0X80000001  222 2001:DB8:0:A1C::/64
   # ADVERTISEMENTS:    6   CHECKSUM TOTAL: 0X000271C6
```

Displaying IPv6 OSPF area link state database

To display the contents of a particular IPv6 OSPF area link state database, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,DATABASE,AREAID=6.6.6.6
EZZ8126I IPV6 OSPF AREA LS DATABASE 829
              ROUTER LSAS (LS TYPE=2001)
LS ORIGINATOR   LS ID      SEQNO       AGE LINKS  RTR-TYPE
63.63.63.63     0          0X80000001  376 1
64.64.64.64     0          0X80000002  321 1      ABR,V
67.67.67.67     0          0X80000004  320 1      ABR,ASBR,V
68.68.68.68     0          0X80000002  595 1
   # ADVERTISEMENTS:    4   CHECKSUM TOTAL: 0X0001D024

              NETWORK LSAS (LS TYPE=2002)
LS ORIGINATOR   LS ID      SEQNO       AGE ROUTERS
68.68.68.68     14         0X80000004  375 4
   # ADVERTISEMENTS:    1   CHECKSUM TOTAL: 0X0000F5CC

              INTER-AREA PREFIX LSAS (LS TYPE=2003)
LS ORIGINATOR   LS ID      SEQNO       AGE PREFIX
64.64.64.64     4          0X80000002  395 2001:DB8:0:108::4/128
64.64.64.64     8          0X80000001  395 2001:DB8:0:108::2/128
64.64.64.64     9          0X80000001  395 2001:DB8:0:10::2/128
64.64.64.64     10         0X80000001  395 2001:DB8:0:10::/64
64.64.64.64     11         0X80000001  395 2:2:2:2:2:2:2:2/128
64.64.64.64     22         0X80000001  375 2001:DB8:0:120::4/128
64.64.64.64     26         0X80000001  321 2001:DB8:0:107::7/128
64.64.64.64     27         0X80000001  321 2001:DB8:0:120::7/128
64.64.64.64     28         0X80000001  321 2001:DB8:0:107::5/128
64.64.64.64     29         0X80000001  321 2001:DB8:0:20::5/128
64.64.64.64     30         0X80000001  321 2001:DB8:0:20::/64
67.67.67.67     15         0X80000002  358 2001:DB8:0:107::7/128
67.67.67.67     16         0X80000001  358 2:2:2:2:2:2:2:2/128
67.67.67.67     19         0X80000001  358 2001:DB8:0:107::5/128
67.67.67.67     20         0X80000001  358 2001:DB8:0:20::5/128
```

```
67.67.67.67     21         0X80000001  358 2001:DB8:0:20::/64
67.67.67.67     25         0X80000001  356 2001:DB8:0:120::7/128
67.67.67.67     26         0X80000001  317 2001:DB8:0:108::4/128
67.67.67.67     27         0X80000001  317 2001:DB8:0:108::2/128
67.67.67.67     28         0X80000001  317 2001:DB8:0:10::2/128
67.67.67.67     29         0X80000001  317 2001:DB8:0:10::/64
67.67.67.67     30         0X80000001  317 2001:DB8:0:120::4/128
   # ADVERTISEMENTS:   22   CHECKSUM TOTAL: 0X000E7320

             LINK LSAS (LS TYPE=0008)
LS ORIGINATOR   LS ID      SEQNO       AGE INTERFACE
63.63.63.63     34         0X80000001  387 NSQDIO1L6
64.64.64.64     16         0X80000001  402 NSQDIO1L6
67.67.67.67     20         0X80000002  640 NSQDIO1L6
68.68.68.68     14         0X80000002  638 NSQDIO1L6
   # ADVERTISEMENTS:    4   CHECKSUM TOTAL: 0X000295E4

             INTRA-AREA PREFIX LSAS (LS TYPE=2009)
LS ORIGINATOR   LS ID      SEQNO       AGE REF-LSTYPE REF-LSID
63.63.63.63     34         0X80000001  387 0X2001     0
63.63.63.63     36         0X80000001  387 0X2001     0
63.63.63.63     38         0X80000001  387 0X2001     0
64.64.64.64     16         0X80000001  402 0X2001     0
64.64.64.64     20         0X80000001  402 0X2001     0
67.67.67.67     20         0X80000002  639 0X2001     0
67.67.67.67     26         0X80000002  639 0X2001     0
68.68.68.68     14         0X80000003  595 0X2001     0
68.68.68.68     16         0X80000001 1738 0X2001     0
68.68.68.68     18         0X80000002  638 0X2001     0
68.68.68.68     65550      0X80000004  375 0X2002     14
   # ADVERTISEMENTS:   11   CHECKSUM TOTAL: 0X00068473
```

Displaying IPv6 OSPF router routes

To display all routes to other routers that have been calculated by IPv6 OSPF and are now present in the routing table, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,ROUTERS
EZZ8125I IPV6 OSPF ROUTERS 820
DEST: 68.68.68.68
  NEXT HOP: FE80::8
  DTYPE:  RTR   RTYPE: SPF    COST: 1        AREA: 6.6.6.6
DEST: 64.64.64.64
  NEXT HOP: FE80::4
  DTYPE:   BR   RTYPE: SPF    COST: 1        AREA: 6.6.6.6
DEST: 65.65.65.65
  NEXT HOP: FE80::5:7
  DTYPE:  RTR   RTYPE: SPF    COST: 1        AREA: 0.0.0.0
DEST: 63.63.63.63
  NEXT HOP: FE80::3
  DTYPE:  RTR   RTYPE: SPF    COST: 1        AREA: 6.6.6.6
DEST: 62.62.62.62
  NEXT HOP: FE80::4
  DTYPE:  RTR   RTYPE: SPF    COST: 2        AREA: 0.0.0.0
DEST: 64.64.64.64
  NEXT HOP: FE80::4
  DTYPE:   BR   RTYPE: SPF    COST: 1        AREA: 0.0.0.0
```

Displaying IPv6 OSPF routing protocol statistics

To display statistics generated by the IPv6 OSPF routing protocol, enter the following command:

```
D TCPIP,TCPCS67,OMP,IPV6OSPF,STATS
EZZ8124I IPV6 OSPF STATISTICS 839
ATTACHED AREAS:                   2  # DIJKSTRA RUNS:                12
OSPF PACKETS RCVD:              619  OSPF PACKETS RCVD W/ERRS:        0
TRANSIT NODES ALLOCATED:         26  TRANSIT NODES FREED:            17
LS ADV. ALLOCATED:              275  LS ADV. FREED:                 175
QUEUE HEADERS ALLOC:             64  QUEUE HEADERS AVAIL:            64
INCREMENTAL SUMM. UPDATES:        5  INCREMENTAL VL UPDATES:          0
INCREMENTAL EXT. UPDATES:        27  PTRS TO INVALID LS ADV:          0
MULTICAST PKTS SENT:            421  UNICAST PKTS SENT:              40
LS ADV. AGED OUT:                 0  LS ADV. FLUSHED:                41
```

Displaying all of the IPv6 RIP information

To display all of the IPv6 RIP information, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,ALL
EZZ8030I IPV6 RIP CONFIGURATION
TRACE6: 2, DEBUG6: 3
STACK AFFINITY: TCPCS4
IPV6 RIP: ENABLED
IPV6 RIP DEFAULT ORIGINATION: DISABLED

EZZ8027I IPV6 RIP INTERFACES
                                  ---------SEND-----------  --RCV--
NAME               MTU STATE IN OUT PRF HST STA DEF RADV PSN  PRF HST
OSAQDIO46         9000    UP  1   0 YES  NO  NO  NO  YES YES  YES  NO


EZZ8031I IPV6 RIP ROUTE ACCEPTANCE
ACCEPT IPV6 RIP UPDATES ALWAYS FOR:
  2001:0DB8:0:A1C::2
  2001:0DB8:0:A1C::1

EZZ8029I GLOBAL IPV6 RIP FILTERS

SEND ONLY: ALL

IGNORE IPV6 RIP UPDATES FROM:
  FE80::1:2:3:8

FILTERS: NORECEIVE       2001:0DB8:0:A1C::/64
```

Displaying information about IPv6 RIP interfaces

To display information about IPv6 RIP interfaces, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,IFS
EZZ8027I IPV6 RIP INTERFACES
                                  ---------SEND-----------  --RCV--
NAME               MTU STATE IN OUT PRF HST STA DEF RADV PSN  PRF HST
OSAQDIO46         9000    UP  1   0 YES  NO  NO  NO  YES YES  YES  NO
```

Displaying information about a specific IPv6 RIP interface

To display information about a specific IPv6 RIP interface, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,IF,NAME=OSAQDIO46
EZZ8028I IPV6 RIP INTERFACE DETAILS
INTERFACE NAME:     OSAQDIO46
INTERFACE ADDRESS:  FE80::1:2:3:1
INTERFACE PREFIX:   RADV 2001:0DB8:1::/48
MTU:                    9000     STATE:                  UP
IN METRIC:              1        OUT METRIC:             0
SEND PREFIX ROUTES:     YES      SEND HOST ROUTES:       NO
SEND STATIC ROUTES:     NO       SEND DEFAULT ROUTES:    NO
SEND RTR. ADV. ROUTES:  YES      SEND POIS. REV. ROUTES: YES
RECEIVE PREFIX ROUTES:  YES      RECEIVE HOST ROUTES:    NO

SEND ONLY:  ALL

FILTERS: NONE
```

Displaying the routes to be unconditionally accepted by IPv6 RIP

To display the routes to be unconditionally accepted by IPv6 RIP, as configured with the IPv6_Accept_RIP_Route statement, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,ACCEPTED
EZZ8031I IPV6 RIP ROUTE ACCEPTANCE
ACCEPT IPV6 RIP UPDATES ALWAYS FOR:
  2001:0DB8:0:A1C::2
  2001:0DB8:0:A1C::1
```

Displaying the global IPv6 RIP filters

To display the global IPv6 RIP filters, enter the following command:

```
D TCPIP,TCPCS4,OMP,IPV6RIP,FILTERS
EZZ8029I GLOBAL IPV6 RIP FILTERS

SEND ONLY: ALL

IGNORE IPV6 RIP UPDATES FROM:
  FE80::1:2:3:8

FILTERS: NORECEIVE       2001:0DB8:0:A1C::/64
```

Displaying the routes in the OMPROUTE IPv6 main routing table

To display all of the routes in the OMPROUTE IPv6 main routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RT6TABLE
EZZ7979I IPV6 ROUTING TABLE 641
DESTINATION: 2:2:2:2:2:2:2:2/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 3:3:3:3:3:3:3:3/128
  NEXT HOP: FE80::3
  TYPE: SPF            COST:  1         AGE: 352
DESTINATION: 4:4:4:4:4:4:4:4/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  1         AGE: 2170
DESTINATION: 5:5:5:5:5:5:5:5/128
  NEXT HOP: FE80::5:7
  TYPE: SPF            COST:  1         AGE: 2197
DESTINATION: 6:6:6:6:6:6:6:6/128
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 7:7:7:7:7:7:7:7/128
  NEXT HOP: ::
  TYPE: SPF *          COST:  0         AGE: 59
DESTINATION: 8:8:8:8:8:8:8:8/128
  NEXT HOP: FE80::8
  TYPE: SPF            COST:  1         AGE: 31
DESTINATION: 2001:DB8:0:10::/64
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  3         AGE: 32
DESTINATION: 2001:DB8:0:10::2/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:30::/60
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 31
DESTINATION: 2001:DB8:0:31::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 31
DESTINATION: 2001:DB8:0:32::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 31
DESTINATION: 2001:DB8:0:33::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:33::3/128
  NEXT HOP: FE80::3
  TYPE: SPF            COST:  1         AGE: 352
DESTINATION: 2001:DB8:0:34::/64
  NEXT HOP: FE80::3 (2)
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:38::8/128
  NEXT HOP: FE80::8
  TYPE: SPF            COST:  1         AGE: 31
DESTINATION: 2001:DB8:0:103::6/128
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:103::7/128
  NEXT HOP: ::
  TYPE: DIR *          COST:  1         AGE: 2209
DESTINATION: 2001:DB8:0:107::5/128
  NEXT HOP: FE80::5:7
  TYPE: SPF            COST:  1         AGE: 2198
```

```
DESTINATION: 2001:DB8:0:107::7/128
  NEXT HOP: ::
  TYPE: SPF *          COST:  0         AGE: 2198
DESTINATION: 2001:DB8:0:108::2/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  2         AGE: 32
DESTINATION: 2001:DB8:0:108::4/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  1         AGE: 32
DESTINATION: 2001:DB8:0:120::/64
  NEXT HOP: ::
  TYPE: SPF *          COST:  1         AGE: 2172
DESTINATION: 2001:DB8:0:120::3/128
  NEXT HOP: FE80::3
  TYPE: SPF            COST:  1         AGE: 352
DESTINATION: 2001:DB8:0:120::4/128
  NEXT HOP: FE80::4
  TYPE: SPF            COST:  1         AGE: 2170
DESTINATION: 2001:DB8:0:120::7/128
  NEXT HOP: ::
  TYPE: SPF *          COST:  0         AGE: 2172
DESTINATION: 2001:DB8:0:120::8/128
  NEXT HOP: FE80::8
  TYPE: SPF            COST:  1         AGE: 31
DESTINATION: 2001:DB8:0:A10::/60
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:A1B::/64
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:A1C::/64
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
DESTINATION: 2001:DB8:0:A1C::6/128
  NEXT HOP: FE80::6:7
  TYPE: RIP            COST:  2         AGE: 0
                       0 NETS DELETED, 5 NETS INACTIVE
```

Displaying the routes to a specific destination in the IPv6 main routing table

To display information about the routes to a specific destination that are in the IPv6 main routing table, enter the following command:

```
D TCPIP,TCPCS4,OMP,RT6TABLE,DEST=2001:0DB8:0:A10::
EZZ7980I IPV6 ROUTE EXPANSION
DESTINATION: 2001:0DB8:0:A10::/60
ROUTE TYPE:  RIP
COST:        2
AGE:         352
NEXT HOP(S): FE80::1:2:3:3                          (OSAQDIO46)
             FE80::1:2:3:4                          (OSAQDIO46)
```

Displaying the routes in all OMPROUTE IPv6 policy-based routing tables

To display all of the routes in all OMPROUTE IPv6 policy-based routing tables, enter the following command:

```
D TCPIP,TCPCS7,OMP,RT6TABLE,PR=ALL
EZZ7979I IPV6 ROUTING TABLE 064
TABLE NAME: SECLOW
DESTINATION: 3:3:3:3:3:3:3:3/128
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  1         AGE: 369
DESTINATION: 2001:DB8:0:30::/60
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  2         AGE: 369
DESTINATION: 2001:DB8:0:31::/64
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  2         AGE: 369
DESTINATION: 2001:DB8:0:32::/64
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  2         AGE: 369
DESTINATION: 2001:DB8:0:33::/64
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  2         AGE: 369
```

```
DESTINATION: 2001:DB8:0:33::3/128
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  1         AGE: 369
DESTINATION: 2001:DB8:0:34::/64
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  2         AGE: 369
DESTINATION: 2001:DB8:0:120::/64
  NEXT HOP: ::
  TYPE:  SPF*          COST:  1         AGE: 377
DESTINATION: 2001:DB8:0:120::3/128
  NEXT HOP: FE80::3
  TYPE:  SPF           COST:  1         AGE: 369
DESTINATION: 2001:DB8:0:120::7/128
  NEXT HOP: ::
  TYPE:  SPF*          COST:  0         AGE: 377
                       0 NETS DELETED
DYNAMIC ROUTING PARAMETERS
  INTERFACE:  NSQDIO1L6        NEXT HOP: FE80::3

TABLE NAME: SECHIGH
DESTINATION: 8:8:8:8:8:8:8:8/128
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  1         AGE: 368
DESTINATION: 2001:DB8:0:30::/60
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 368
DESTINATION: 2001:DB8:0:31::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 368
DESTINATION: 2001:DB8:0:32::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 368
DESTINATION: 2001:DB8:0:33::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 368
DESTINATION: 2001:DB8:0:34::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 368
DESTINATION: 2001:DB8:0:38::8/128
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  1         AGE: 368
DESTINATION: 2001:DB8:0:120::/64
  NEXT HOP: ::
  TYPE:  SPF*          COST:  1         AGE: 376
DESTINATION: 2001:DB8:0:120::7/128
  NEXT HOP: ::
  TYPE:  SPF*          COST:  0         AGE: 376
DESTINATION: 2001:DB8:0:120::8/128
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  1         AGE: 368
                       0 NETS DELETED
DYNAMIC ROUTING PARAMETERS
  INTERFACE:  NSQDIO1L6        NEXT HOP: FE80::8
```

Result: If a policy-based route table is configured with no IPv6 dynamic routing parameters, OMPROUTE has no knowledge of that route table for IPv6. The route table is not included in this display. If no policy-based route tables are configured with IPv6 dynamic routing parameters, message EZZ8150I is issued.

Displaying the routes in an OMPROUTE IPv6 policy-based routing table

To display all of the routes in an OMPROUTE IPv6 policy-based routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RT6TABLE,PR=SECHIGH
EZZ7979I IPV6 ROUTING TABLE 070
TABLE NAME: SECHIGH
DESTINATION: 8:8:8:8:8:8:8:8/128
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  1         AGE: 574
DESTINATION: 2001:DB8:0:30::/60
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 574
DESTINATION: 2001:DB8:0:31::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 574
DESTINATION: 2001:DB8:0:32::/64
```

```
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 574
DESTINATION: 2001:DB8:0:33::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 574
DESTINATION: 2001:DB8:0:34::/64
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  2         AGE: 574
DESTINATION: 2001:DB8:0:38::8/128
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  1         AGE: 574
DESTINATION: 2001:DB8:0:120::/64
  NEXT HOP: ::
  TYPE:  SPF*          COST:  1         AGE: 582
DESTINATION: 2001:DB8:0:120::7/128
  NEXT HOP: ::
  TYPE:  SPF*          COST:  0         AGE: 582
DESTINATION: 2001:DB8:0:120::8/128
  NEXT HOP: FE80::8
  TYPE:  SPF           COST:  1         AGE: 574
                       0 NETS DELETED
DYNAMIC ROUTING PARAMETERS
  INTERFACE:  NSQDIO1L6        NEXT HOP: FE80::8
```

Result: If the policy-based route table is configured with no IPv6 dynamic routing parameters, OMPROUTE has no knowledge of the route table for IPv6. If no policy-based route tables are configured with IPv6 dynamic routing parameters, message EZZ8150I is issued.

Displaying the routes to a specific destination in an IPv6 policy-based routing table

To display information about the routes to a specific destination that are in an IPv6 policy-based routing table, enter the following command:

```
D TCPIP,TCPCS7,OMP,RT6TABLE,PR=SECHIGH,DEST=2001:DB8:0:34::/64
EZZ7980I IPV6 ROUTE EXPANSION 088
TABLE NAME: SECHIGH
DESTINATION: 2001:DB8:0:34::/64
ROUTE TYPE:  SPF
COST:        2
AGE:         717
NEXT HOP(S): FE80::8                                (NSQDIO1L6)
```

Result: If the policy-based route table is configured with no IPv6 dynamic routing parameters, OMPROUTE has no knowledge of the route table for IPv6. If no policy-based route tables are configured with IPv6 dynamic routing parameters, message EZZ8150I is issued.

Displaying all of the IPv6 generic information

To display all of the IPv6 generic information, enter the following command:

```
D TCPIP,TCPCS4,GENERIC6,ALL
EZZ8053I IPV6 GENERIC CONFIGURATION
TRACE6: 2, DEBUG6: 3
IPV6 TRACE DESTINATION: /TMP/OMPROUT6.DBG
STACK AFFINITY: TCPCS4

EZZ8060I IPV6 GENERIC INTERFACES
NAME               MTU STATE CONFIGURED
VIPA16           65535    UP     YES
```

Displaying information about IPv6 generic interfaces

To display information about IPv6 generic interfaces (that is, interfaces defined in the OMPROUTE configuration file using IPv6_INTERFACE statements, or IPv6 interfaces not defined to OMPROUTE but learned from the stack), enter the following command:

```
D TCPIP,TCPCS4,OMP,GENERIC6,IFS
EZZ8060I IPV6 GENERIC INTERFACES
```

```
NAME              MTU STATE CONFIGURED
VIPA16          65535    UP     YES
```

Displaying information about a specific IPv6 generic interface

To display information about a specific IPv6 generic interface, enter the following command:

```
D TCPIP,TCPCS4,OMP,GENERIC6,IF,NAME=VIPA16
EZZ8065I IPV6 GEN INTERFACE DETAILS
INTERFACE NAME:     VIPA16
INTERFACE ADDRESS:  2001:0DB8:6:6:6:6:6:6
INTERFACE PREFIX:   STAT 2001:0DB8:6::/48
MTU:               65535
STATE:              UP
CONFIGURED:         YES
```

Sample OMPROUTE configuration files

The following example shows an OSPF and IPv6 RIP environment (from TCPCS4 in the Figure 48 on page 282).

```
;*********************************
; OSPF Configuration Statements   *
;*********************************
OSPF
     RouterID=4.4.4.4;
Area
     Area_Number = 0.0.0.0;
Area
     Area_Number = 1.1.1.1;
OSPF_Interface
     IP_Address=9.67.108.4
     Name = CTC4TO2
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=0.0.0.0
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.106.4
     Name = CTC4TO7
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.105.4
     Name = CTC4TO8
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.101.4
     Name = CTC4TO3
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=4.4.4.4
     Name = VIPA1A
     Subnet_Mask=255.255.255.252
     Attaches_To_Area=1.1.1.1
     Cost0 = 1;
Virtual_Link
     Virtual_Endpoint_RouterID=7.7.7.7
     Links_Transit_Area=1.1.1.1;
;*********************************
; IPv6 RIP Configuration Statements *
;*********************************
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1C::1;
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1C::2;
IPv6_RIP_Filter=(noreceive,2001:0DB8:0:A1C::/64);
```

```
IPv6_RIP_Interface
     Name = OSAQDIO46;
IPv6_Default_Route
     Name=OSAQDIO46
     Next_Hop=FE80::1:2:3:4;
```

The following example shows mixed OSPF, IPv4 RIP, and IPv6 RIP environments (from TCPCS7 in Figure 48 on page 282).

```
;*********************************
;  OSPF Configuration Statements  *
;*********************************
OSPF
     RouterID=7.7.7.7;
Area
     Area_Number = 0.0.0.0;
Area
     Area_Number = 1.1.1.1;
AS_Boundary_Routing
     Import_Subnet_Routes=YES
     Import_RIP_Routes=YES;
OSPF_Interface
     IP_Address=9.67.107.7
     Name = CTC7TO5
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=0.0.0.0
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.106.7
     Name = CTC7TO4
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.102.7
     Name = CTC7TO3
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.100.7
     Name = CTC7TO8
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     MTU = 1024
     Cost0 = 1;
OSPF_Interface
     IP_Address=9.67.104.7
     Name = NBMA7
     Subnet_Mask=255.255.255.0
     Attaches_To_Area=1.1.1.1
     Non_Broadcast=YES
     NB_Poll_Interval=180
     MTU = 1024
     Cost0 = 1
     DR_Neighbor=9.67.104.15
     No_DR_Neighbor=9.67.104.16
     No_DR_Neighbor=9.67.104.25;
OSPF_Interface
     IP_Address=7.7.7.7
     Name = VIPA1A
     Subnet_Mask=255.255.255.252
     Attaches_To_Area=1.1.1.1
     Cost0 = 1;
Range
     IP_Address=9.67.101.0
     Subnet_Mask=255.255.255.0
     Area_Number=1.1.1.1
     Advertise=NO;
Virtual_Link
     Virtual_Endpoint_RouterID=4.4.4.4
     Links_Transit_Area=1.1.1.1;
;*********************************
;  RIP Configuration Statements  *
;*********************************
Originate_RIP_Default
     Condition=Always;
```

```
Accept_RIP_Route
     IP_Address=30.1.1.4;
Accept_RIP_Route
     IP_Address=30.1.1.8;
Filter=(nosend,10.1.1.0,255.255.255.0);
RIP_Interface
     IP_Address=9.67.103.7
     Name = CTC7TO6
     Subnet_Mask=255.255.255.0
     Receive_Dynamic_Hosts=NO
     MTU = 1024
     RipV2=YES;
;************************************
; IPv6 RIP Configuration Statements *
;************************************
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1B::1;
IPv6_Accept_RIP_Route
     IP_Address=2001:0DB8:0:A1B::2;
IPv6_RIP_Filter=(noreceive,2001:0DB8:0:A1B::/64);
IPv6_RIP_Interface
     Name = OSAQDIO76;
```

The following example shows a pure RIP environment (from TCPCS6 in Figure 48 on page 282).

```
;*********************************
;  RIP Configuration Statements  *
;*********************************
RIP_Interface
     IP_Address=9.67.103.6
     Name = CTC6TO7
     Subnet_Mask=255.255.255.0
     MTU = 1024
     Send_Static_Routes=YES
     Send_Host_Routes=YES
     RipV2=YES;
Interface
     IP_Address=6.6.6.6
     Name = VIPA1A
     Subnet_Mask=255.255.255.252;
```

The following example shows a pure IPv6 OSPF environment (from TCPCS64 in Figure 49 on page 283).

```
;************************************
; IPv6 OSPF Configuration Statements *
;************************************
IPv6_OSPF
     RouterID = 64.64.64.64;
IPv6_Area
     Area_Number = 0.0.0.0;
IPv6_Area
     Area_Number = 6.6.6.6;
IPv6_OSPF_Interface
     Name = NSQDIO4L6
     Prefix = 2001:0DB8:0:120::/64
     Attaches_to_Area = 6.6.6.6;
IPv6_OSPF_Interface
     Name = VIPA1A6
     Attaches_to_Area = 6.6.6.6
     Cost = 1;
IPv6_OSPF_Interface
     Name = MPCPTP4TO2
     Attaches_to_Area = 0.0.0.0
     Cost = 1;
IPv6_Virtual_Link
     Virtual_Endpoint_RouterID = 67.67.67.67
     Links_Transit_Area = 6.6.6.6;
```

Policy-based routing

Policy-based routing enables the TCP/IP stack to make routing decisions that take into account criteria other than just the destination IP address. The additional criteria can include job name, source port, destination port, protocol type (TCP or UDP), source IP address, NetAccess security zone, and multilevel secure environment security label. Policy-based routing might be useful in the following sample scenarios:

- You might want to favor high-bandwidth links for batch traffic, but for interactive traffic you prefer low-latency links. If so, you could define policies such that Telnet traffic is routed over the low-latency links, and FTP traffic is routed over the high-bandwidth links.
- You could define a policy to ensure that traffic that is tagged with a security label and zone is routed to a secured network over an appropriate outbound interface.
- You might want to control the links that are used by Enterprise Extender traffic to keep that traffic from being impacted by other IP traffic loads.

Restrictions:

- Policy-based routing applies only to TCP and UDP traffic that originates at the TCP/IP stack. The following types of traffic are routed always by using the main route table, even when policy-based routing is in use:
  - Traffic that uses protocols other than TCP and UDP
  - Traffic that the TCP/IP stack uses
  - Traffic for the Ping and Traceroute commands
- If Common INET (CINET) is used to run multiple z/OS Communications Server TCP/IP stacks concurrently, CINET has no knowledge of the policy-based route tables that are used by those TCP/IP stacks. CINET has knowledge only of the routes in the main route table of each TCP/IP stack. Avoid use of policy-based routing in a CINET environment, unless at least one of the following conditions is true:
  - All applications establish affinity with a particular TCP/IP stack.
  - The route destinations in each TCP/IP stack route table are mutually exclusive with the route destinations on the other TCP/IP stacks, including the default route.

Options for configuring policy-based routing

Policy-based routing is configured using a set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the policy-based routing for each TCP/IP stack. In a complex environment, this file can become large. For this reason, there are two alternatives for creating the Policy Agent files.

Option 1: Use the IBM Configuration Assistant for z/OS Communications Server

The IBM Configuration Assistant for z/OS Communications Server, an optional GUI-based tool, provides a guided interface to simplify configuration of TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the Policy Agent files.

The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

Through a series of wizards and online help panels, you can use the IBM Configuration Assistant for z/OS Communications Server to create policy-based routing configuration files for any number of z/OS images with any number of TCP/IP stacks per image.

With the IBM Configuration Assistant for z/OS Communications Server you use traffic descriptors, reusable objects that define traffic by its ports, protocol, security zone, security label, or sending application job name. The IBM Configuration Assistant for z/OS Communications Server comes with a number of IBM-supplied traffic descriptors that are easily applied, or you can use the IBM-supplied definitions as the basis for your own set of reusable objects. For each TCP/IP stack, you then create a set of policy-based route tables and routing rules, which map traffic descriptors and IP addresses to the policy-based route tables to be used when making routing decisions for that traffic.

The IBM Configuration Assistant for z/OS Communications Server can dramatically reduce the amount of time that is required to create policy-based routing files, contributing to ease of configuration and

maintenance. Using the GUI ensures that you have a consistent and easily manageable interface for implementing policy-based routing.

This information primarily describes option 2, manual configuration. However, if you are using the IBM Configuration Assistant for z/OS Communications Server, reading this information will help you understand policy-based routing concepts and the relationship between Policy Agent and policy-based routing.

Option 2: Manual configuration

You can manually create the policy-based routing configuration files by coding all the required statements in a z/OS UNIX file or MVS data set. This information describes the procedure for creating a routing policy by manually creating and editing the configuration files. For details about the routing policy statements, see z/OS Communications Server: IP Configuration Reference.

Specifying the routing configuration file based on Policy Agent role

The Policy Agent can act as a policy server, a policy client, or neither. For more information about these different roles, see "Policy types and infrastructure overview" on page 787. Regardless of which option you use to configure Routing policies, the resulting configuration files need to be specified using different statements, depending on the role of the Policy Agent.

- If you are using the Policy Agent as a policy client that retrieves Routing policies from the policy server, specify the configuration files using the DynamicConfigPolicyLoad statement on the policy server.
- If you are using the Policy Agent as a policy client, but the policy client does not retrieve Routing policies from the policy server, specify the configuration files using the RoutingConfig statement on the policy client.
- If you are not using a policy client and policy server environment, specify the configuration files using the RoutingConfig statement on the single Policy Agent.

When specifying configuration files, keep in mind where the files should exist, based on the role of the Policy Agent.

Routing policy configuration

Routing policy is provided to the stack by the Policy Agent. The Policy Agent main configuration file contains a TcpImage statement for each stack that is to receive policy, and can optionally contain a CommonRoutingConfig statement that identifies a local shared routing policy file. The TcpImage statement identifies the z/OS UNIX file or MVS data set that contains policy for that stack. This policy file can contain a RoutingConfig statement to identify the z/OS UNIX file or MVS data set that contains the local routing policy. The RoutingConfig statement is required for each stack that is to receive routing policy. If both a RoutingConfig statement and a CommonRoutingConfig statement are defined, the specified CommonRoutingConfig file is processed before the RoutingConfig policy file specified for that stack.

On the policy server, use the DynamicConfigPolicyLoad statement to specify the remote routing policies. On the policy client, use the PolicyServer statement to retrieve the remote routing policies from the policy server.

In the routing policy file, routing rules define sets of conditions that are compared when a route is being selected to send TCP or UDP traffic that originates in the TCP/IP stack. If a rule match is found, an appropriate route for the traffic is selected based on the route tables that are specified in the action that is associated with the rule.

Routing rules

A RoutingRule statement consists of a set of conditions that are compared against the traffic that is being sent. When a match is found, policy lookup stops and the traffic is assigned the actions that are associated with the rule. The rule conditions are as follows:

IPSourceAddr
Source IP address or addresses. The source IP address for a TCP outbound connection, or for a UDP outbound packet, can be influenced by a number of configuration and application options. For the hierarchy of ways that the source IP address of an outbound packet is determined, see "Source IP address selection" on page 244. For the following source IP address selection methods, a route lookup is needed to determine the source IP address.

- SOURCEVIPA: Static VIPA address from the HOME list (IPv4 interface defined with the LINK statement) or from the SOURCEVIPAINTERFACE parameter (IPv4 or IPv6 interface defined with the INTERFACE statement)
- HOME IP address of the interface over which the packet is sent

Do not use the IpSourceAddr condition as a selector for traffic that relies on these methods to select its source IP address. At the time that the route lookup is performed, the source IP address is not yet selected.

IPDestAddr
Destination IP address or addresses.

SourcePortRange
Source port or ports.

DestinationPortRange
Destination port or ports.

Protocol
TCP or UDP.

Jobname
Job name of the sending application or wildcard job name.

SecurityZone
NetAccess security zone that outbound traffic must match. The outbound traffic destination IP address is used to determine the NetAccess security zone in the NetAccess table that is defined in the TCP/IP profile. For more information about network access control and the NETACCESS TCP/IP profile statement, see z/OS Communications Server: IP Configuration Reference.

SecurityLabel
Multilevel secure networking security label of the NetAccess security zone that outbound traffic must match. The outbound traffic destination IP address is used to determine the NetAccess security zone of the packet in the NetAccess table that is defined in the TCP/IP profile. The security label is the label that is associated with the NetAccess zone. For more information, see Chapter 4, "Preparing for IP networking in a multilevel secure environment," on page 191.

If a condition is not specified, that condition is not considered when the rule and the traffic are compared for a match. You can specify multiple values for the conditions, either directly in the condition or as a referenced group.

Each RoutingRule statement can also have a priority. Priority values can be integers in the range 1 – 2000000000; 2000000000 is the highest priority. When assigning priorities, skip some values to accommodate future rule insertion between existing rules.

If traffic does not map to any of the active routing rules, the IP layer routes traffic by searching the main route table.

Tip: If traffic can map to more than one rule, always use priority and leave priority space between rules.

A RoutingRule statement must reference an action by using the RoutingActionRef parameter. The RoutingActionRef parameter includes the name of a globally defined RoutingAction statement.

Routing actions

The RoutingAction statement must specify a minimum of one policy-based route table using the RouteTableRef parameter, or it must specify UseMainRouteTable YES. Specify UseMainRouteTable YES and no RouteTableRef parameters if you want traffic that matches the corresponding RoutingRule statement to be routed using only the main route table.

A maximum of eight RouteTableRef parameters can be specified on a RoutingAction statement. Each RouteTableRef parameter includes the name of a globally defined RouteTable statement. The order in which the RouteTableRef parameters are specified on the RoutingAction statement determines the order in which the TCP/IP stack uses the route tables when making routing decisions. The stack searches the first table for a route that can be used to reach the destination of the traffic. If no usable route (host, subnet, network, or default) exists in that route table, the next table is searched. This continues until a usable route to the destination is found, or all specified tables are exhausted. Specify the UseMainRouteTable parameter to indicate whether the main route table should also be searched when no usable route is located in any of the specified policy-based route tables. For a description of the algorithm used by the IP layer to search a route table for a usable route to a destination, see "Route selection algorithm" on page 281.

Performance guideline: There is a performance cost for each policy-based route table that is searched on a route lookup. Minimize the number of RouteTableRef parameters that you specify on a RoutingAction statement.

Routing tables

The RouteTable statement can include the following items:

- A set of static routes (both replaceable and non-replaceable static routes are supported)
- A set of dynamic routing parameters that are used for the following purposes:
  - To control the scope of dynamic routes that OMPROUTE computes
  - For IPv6 only, to control the scope of router advertisement routes added to the table
- A dynamic XCF routes indicator

A static route is included by using the Route parameter of the RouteTable statement. A dynamic routing parameter is included by using the DynamicRoutingParms parameter of the RouteTable statement. A dynamic routing parameter consists of an interface, and optionally a next-hop router.

OMPROUTE adds dynamic routes to a policy-based route table only if the interfaces that the routes use are included in the dynamic routing parameters for that table. If a next-hop router is specified with an interface, dynamic indirect routes that use the specified interface are added only if they also use the specified next-hop router.

For IPv6, router advertisement routes are added to a policy-based route table only if the interface that received the router advertisement is included in the dynamic routing parameters for that table. If a next-hop router is specified with an interface, indirect router advertisement routes that use the specified interface are added only if the router advertisement was sent by the specified next-hop router.

Including only static routes, only dynamic routing parameters, or both in a RouteTable statement produces the following results:

- If only static routes are included, the static routes are added to the policy-based route table, but no dynamic routes are added to the table by OMPROUTE and no IPv6 router advertisement routes are added to the table.
- If only dynamic routing parameters are included, dynamic routes are added to the policy-based route table by OMPROUTE and IPv6 router advertisement routes are added to the table based on received router advertisements, but no static routes are added to the table.
- If both static routes and dynamic routing parameters are included, the static routes are added to the policy-based route table. OMPROUTE updates the table with the appropriate dynamic routes, and IPv6 router advertisement routes are added to the table based on received router advertisements.

A dynamic route that is added to a policy-based route table by OMPROUTE is the best route that is known by OMPROUTE to that destination. The calculation of the best route is based on the current network topology. The best route might be a route that is learned from the OSPF protocol or from the RIP protocol, depending on the OMPROUTE configuration. For a description of these protocols, see "IPv4 dynamic routing using OMPROUTE" on page 288 and "IPv6 dynamic routing using OMPROUTE" on page 291. When OMPROUTE is computing the best routes for a policy-based route table, it considers only routes that adhere to the dynamic routing parameters configured on the RouteTable statement as described.

Performance guideline: Each policy-based route table that is configured for dynamic routing adds more processing cost to OMPROUTE. Do not configure duplicate route tables, and limit the use of policy-based route tables that use dynamic routing.

If DynamicXCFRoutes Yes is specified on the RouteTable statement and IPCONFIG DYNAMICXCF is specified in the TCP/IP profile, direct host routes to IPv4 dynamic XCF addresses on other TCP/IP stacks are added to the route table when the IPv4 dynamic XCF links to those stacks are active. If DynamicXCFRoutes6 Yes is specified on the RouteTable statement and IPCONFIG6 DYNAMICXCF is specified in the TCP/IP profile, direct host routes to IPv6 dynamic XCF addresses on other TCP/IP stacks are added to the route table when the IPv6 dynamic XCF links to those stacks are active. These same routes are automatically generated in the main route table when dynamic XCF links are active. For information about the dynamic XCF function and the definitions that are automatically generated when IPCONFIG DYNAMICXCF or IPCONFIG6 DYNAMICXCF is specified in the TCP/IP profile, see "Dynamic XCF" on page 440.

After a new routing policy is installed in the TCP/IP stack, traffic is mapped by using the new policy, even for existing connections.

Getting started with policy-based routing

Assume you have a TCP application that has the name SECBATCH running on a z/OS TCP/IP stack. Because this application handles sensitive data sent as batch traffic, you want the traffic to be sent only over secure networks and you want to prefer high bandwidth links over lower bandwidth links. The application creates a TCP socket bound to IP address INADDR_ANY and port 7000.

For the tasks that must be completed to configure the local routing policy to control routing decisions made for traffic sent by the application, see "Configuring policy-based routing" on page 354. Assuming that the link names specified on the routing policy statements are already defined to TCP/IP and OMPROUTE, you do not need to perform additional configuration in the TCP/IP profile or in the OMPROUTE configuration file to enable policy-based routing.

Configuring policy-based routing

Before you begin

This topic describes the tasks that must be completed to configure the local routing policy to control routing decisions made for traffic sent by an application. Assuming that the link names specified on the routing policy statements are already defined to TCP/IP and OMPROUTE, you do not need to perform additional configuration in the TCP/IP profile or in the OMPROUTE configuration file to enable policy-based routing.

Procedure

1. Determine your requirements for the TCP/IP stack to make routing decisions based on more than just destination IP address.

   The additional criteria can include job name, source port, destination port, protocol type (TCP or UDP), source IP address, NetAccess security zone, and security label.
2. Create Policy Agent files
   a) Create a Policy Agent main configuration file containing a TcpImage statement for the stack.
   b) Create a Policy Agent image configuration file for the stack.
   c) If routing policies are to be retrieved from the policy server, create image-specific routing configuration files, and optionally, common routing configuration files, on the policy server.
3. Add routing configuration
   a) For local Routing policies, add a RoutingConfig statement to the Policy Agent image configuration file, identifying the RoutingConfig policy file location:

```
RoutingConfig    configFilepath
```

b) For remote Routing policies, add a PolicyServer statement to the policy client image configuration file:

```
PolicyServer
{
   ClientName  name
   PolicyType  Routing
   {
      …
   }
   …
}
```

Add a DynamicConfigPolicyLoad statement to the policy server main configuration file:

```
DynamicConfigPolicyLoad  clientname
{
   PolicyType Routing
   {
      PolicyLoad  configFilepath
   }
   …
}
```

4. Add statements to the Routing policy file to configure the policy-based route tables to be used by the TCP/IP stack for routing the application traffic.

   Add the following Routing policy statements to the *configFilepath* file:

```
RouteTable            SecFast    # Secure link, high bandwidth
{
  #   Static Routes:
  #     Destination Subnet Mask  First Hop   Link Name Packet Size Options
  Route DEFAULT                  9.67.101.3  SECHIGH1  MTU 2000    Replaceable
  #
  #   Dynamic Routing Parameters:
  #                      Link Name    First Hop
  DynamicRoutingParms   SECHIGH2
  DynamicRoutingParms   SECHIGH1     9.67.101.3
}

RouteTable            SecSlow     # Secure link, low bandwidth
{
  #   Static Routes:
  #     Destination Subnet Mask  First Hop   Link Name Packet Size Options
  Route DEFAULT                  9.67.106.7  SECLOW1   MTU 2000    Replaceable
  #
  #   Dynamic Routing Parameters:
  #                      Link Name    First Hop
  DynamicRoutingParms   SECLOW2      9.67.104.3
  DynamicRoutingParms   SECLOW1      9.67.106.7
  DynamicRoutingParms   SECLOW1      9.67.106.15
}
```

   Table SecFast contains a replaceable static default route. The dynamic routing parameters for SecFast direct OMPROUTE to compute routes that use only link SECHIGH2, with any first hop, and link SECHIGH1, with a first hop of 9.67.101.3.

   Table SecSlow contains a replaceable static default route. The dynamic routing parameters for SecSlow direct OMPROUTE to compute only routes that use link SECLOW2, with a first hop of 9.67.104.3, and link SECLOW1, with a first hop of either 9.67.106.7 or 9.67.106.15.

5. Add statements to the Routing policy file to ensure that the application traffic is sent over only secure links, favoring high bandwidth links over lower bandwidth links.

   Add the following Routing policy statements to the *configFilepath* file:

```
RoutingRule                   SecBatchRule
{
  TrafficDescriptor
    {
      Protocol                     TCP
```

```
      SourcePortRange                 7000
      Jobname                         SECBATCH
    }
  RoutingActionRef                    SecBatchAction
}
RoutingAction                    SecBatchAction
{
  UseMainRouteTable                   No
  RouteTableRef                       SecFast
  RouteTableRef                       SecSlow
}
```

6. Start Policy Agent

Results

You know you are done when the Routing policies are installed to the TCP/IP stack and the following console message is displayed:

```
EZZ8771I PAGENT CONFIG POLICY PROCESSING COMPLETE FOR  image : ROUTING
```

Considerations for using policy-based routing with IP security

Policy-based routing allows traffic that is described in a routing rule to be routed by using one or more route tables. When IP security is active on a TCP/IP stack that is using policy-based routing, it is important to understand how the two functions interact. On a stack with IP security active, traffic can be encapsulated in an AH, ESP, or UDP-encapsulated ESP header. An additional IP header can be added if the encapsulated traffic is being sent to a security gateway (that is, the remote tunnel endpoint is not the same as the remote data endpoint). A matching routing rule is selected based on the characteristics of the original non-encapsulated traffic. The route tables that are associated with the matching routing rule and action are used to route the encapsulated traffic.

For example, assume the following configuration:

- An IPSec filter rule, FilterRule1, is configured for traffic with source address 9.1.1.1 and destination address 167.0.0.0/8 to have IPSec protection. Traffic is encapsulated and sent to router 9.2.2.2, the security gateway.
- A routing rule, FTPRULE, is configured for FTP traffic with source address 9.1.1.1. The associated action specifies that route table FTPRTES is to be used to route traffic and that the main route table is not to be searched.

Given this configuration, the following processing is performed for FTP traffic sent from IP address 9.1.1.1 to IP address 167.1.1.1:

1. The FTP traffic matches routing rule FTPRULE, and a route is found in route table FTPRTES that is used to route to destination 167.1.1.1.
2. The FTP traffic matches IPSec filter rule FilterRule1.
3. The FTP traffic is encapsulated and a new IP header is added with destination address 9.2.2.2.
4. The encapsulated packet is routed based on the routes that are defined in route table FTPRTES. To successfully send the traffic, route table FTPRTES must also contain a route that is used to route to destination 9.2.2.2. Otherwise, the traffic would be sent by using the route that is found to destination 167.1.1.1. The success of the traffic depends on network connectivity.

Requirement: When a routing rule applies to traffic that is to be IPSec encapsulated and sent to a security gateway, the route tables that are associated with the routing rule and action must contain a route that can be used to reach the security gateway, as well as a route that can be used to reach the original destination.

For more information about IP security, see Chapter 17, “IP security,” on page 883.

Considerations for mixed routing environments

Read this topic if you will be using any combination of static, dynamic, and policy-based routing on a single TCP/IP stack. These considerations will help you understand how these different types of routing interact, and how to configure them to get the results that you want.

Use of static routing with OMPROUTE

Do not use non-replaceable static routes with OMPROUTE because those routes cannot be dynamically updated in response to network topology changes. An exception is when routes must be defined to destinations that, for some reason, cannot be learned dynamically through the routing protocol. Use the BEGINROUTES statement in PROFILE.TCPIP to define IPv4 or IPv6 static routes if they are required.

TCP/IP treats static routes that are defined as replaceable on the BEGINROUTES statement as last-resort routes. If a dynamic route is learned for the destination that was specified in a replaceable static route, the dynamic route replaces the static route in the route table. Additionally, if a replaceable static route is replaced with a dynamic route, TCP/IP always retains knowledge of the replaceable static route and reinstalls it if the destination becomes unreachable using dynamic routes. TCP/IP does not relearn replaceable static routes that were replaced. For this reason, replaceable static routes can be used with OMPROUTE as backup routes to use if nothing is found dynamically.

Another situation in which static routes might be required is when multiple, equal-cost routes to a destination are needed and the RIP or IPv6 RIP routing protocol is used. Static routes might be required in this case because, except for directly attached resources, the RIP and IPv6 RIP protocols do not create multiple, equal-cost routes to a destination. In other words, if multiple adjacent routers are advertising through RIP or IPv6 RIP that they can reach the same destination, OMPROUTE adds a route to the TCP/IP route table through only one of those adjacent routers. If more than one of these routes must exist, they must be statically configured in PROFILE.TCPIP by using the BEGINROUTES statement. For example, in Figure 48 on page 282, this configuration would be necessary if you wanted host TCPCS4 to have two routes to the IPv6 prefix 2001:DB8:0:A10::/60 (one through router A and one through router B).

If a TCP/IP stack has multiple interfaces to a directly attached network and you want to use one interface for input packets and one for output packets (traffic splitting), you can use static routes. To use traffic splitting, you can define a static route for only one interface, forcing all output packets to use that interface. The other routers on the directly attached network must be defined with a similar static route, but for the other interface. Although this method is the easiest way to implement traffic splitting, if one of the interfaces fails, a host might become unreachable even though the other interface remains active.

Tip: A more robust way of accomplishing traffic splitting is to use dynamic routes and make one route preferred over the other through the configured interface costs. For more information, see step “7” on page 319.

The BSDROUTINGPARMS statement in PROFILE.TCPIP is not used when the OMPROUTE routing daemon is used. Instead, the IPv4 interface characteristics, including subnet mask, are defined in the OMPROUTE configuration file.

Use of IPv6 static routing with router advertisements

Do not use non-replaceable static routes with IPv6 router discovery, when those routes are to destinations that are to be learned through received router advertisements. If these non-replaceable static routes are defined, they cannot be dynamically updated in response to network topology changes.

TCP/IP treats replaceable static routes as last-resort routes. These routes can be replaced by dynamic (router discovery) routes. In addition, if a replaceable static route is replaced with a dynamic route, TCP/IP always retains knowledge of the replaceable static route and can reinstall it if the destination becomes unreachable using dynamic routes. TCP/IP does not need to relearn replaceable static routes that are replaced. For this reason, replaceable static routes can be used with IPv6 router discovery as backup routes, to be used if nothing is learned dynamically.

Use of policy-based routing with static or dynamic routing

If a policy-based route table is configured with both static routes and dynamic routing parameters, review the considerations in "Use of static routing with OMPROUTE" on page 357 and in "Use of IPv6 static routing with router advertisements" on page 357. The same considerations apply with one exception; references that are made to static routes in those topics refer to static routes that are configured by using the Route parameter on the RouteTable policy statement.

When policy-based routing is used, policies are defined that instruct the TCP/IP stack to use specific policy-based route tables when it is making routing decisions, which are based on the traffic that is being routed. The main route table is used only when one of the following conditions are true:

- The traffic for which a route is being selected does not match any active policy-based routing rule.
- The traffic for which a route is being selected matches an active policy-based routing rule. However, no usable route exists in the policy-based route tables that are specified by policy to be used for the traffic, and the policy indicates that the main route table is to be used when this situation occurs.

Verifying static, dynamic, and policy-based routing

- If OMPROUTE is used for the OSPF protocol only and AUTOLOG is not configured correctly (see step "2" on page 297), OMPROUTE is periodically restarted and the following messages are displayed:

```
$HASP100 OMPROUTE ON STCINRDR
$HASP373 OMPROUTE STARTED
IEF403I OMPROUT1 - STARTED
      OMPROUT1    OMPROUTE    BPXBATCH    0000
EZZ7800I OMPROUTE STARTING
EZZ7872I OMPROUTE FOUND ANOTHER ROUTING APPLICATION ALREADY
ACTIVE
EZZ8074I OMPROUTE PROCESSING ERROR
EZZ7805I OMPROUTE EXITING ABNORMALLY - RC(11)
OMPROUT1    *OMVSEX     BPXPRECP    0011
IEF404I OMPROUT1 - ENDED
$HASP395 OMPROUT1 ENDED
```

- If a configuration statement in the OMPROUTE configuration file has a missing semicolon, the syntax checker might issue the following message:

```
EZZ7830I SYNTAX ERROR AT LINE 22 OF OMPROUTE CONFIGURATION FILE
PROCESSING END OF FILE
```

- If policy-based routing is in use, see the Netstat ALL/-A command report to determine the policy-based routing information in use by a TCP connection or UDP socket. If the value Yes is displayed for RoutingPolicy, then the RoutingTableName and RoutingRuleName values reflect the last traffic attempted to be sent by the displayed TCP connection or UDP socket. RoutingRuleName displays the name of the routing policy rule that was last mapped to the traffic sent by the displayed TCP connection or UDP socket. RoutingTableName displays the name of the route table that was used to select the route for the last traffic that was sent by the displayed TCP connection or UDP socket. This value can be the name of a policy-based route table that is associated with the displayed RoutingPolicyRule, EZBMAIN if the main route table was used, or *NONE* if a route was not found.

  If no traffic was sent or the last traffic sent did not map to a configured routing rule, the value No is displayed for RoutingPolicy. RoutingTableName and RoutingRuleName are not included in the display when RoutingPolicy has the value No.

Verifying connections with Netstat, Ping, and Traceroute

The interfaces were verified with the instructions in Chapter 2, "IP configuration overview," on page 11. The first thing to verify is that the devices and interfaces are started. In the case of point-to-point links like the CTCs in TCPCS4, the following message is written to the z/OS console when the device starts:

```
EZZ4313I INITIALIZATION COMPLETE FOR DEVICE CTCE02
```

In the case of IPv6 interfaces like OSAQDIO46 in TCPCS4, the following message is written to the z/OS console when the interface starts:

```
EZZ4340I INITIALIZATION COMPLETE FOR INTERFACE OSAQDIO46
```

The same information can be determined by using the Netstat DEvlinks/-d command. The following example is a portion of the output of the Netstat DEvlinks/-d command with the CTCE02 device shown as ready. The Netstat DEvlinks/-d command can be issued on TCPCS4 and TCPCS7 to verify that the devices on both systems are ready.

```
DevName: CTCE02            DevType: CTC       DevNum: 0E00
  DevStatus: Ready
  LnkName: CTC4TO7           LnkType: CTC         LnkStatus: Ready
    NetNum: 0    QueSize: n/a
    ActMtu: 32760
  Routing Parameters:
    MTU Size: 01500             Metric: 01
    DestAddr: 0.0.0.0           SubnetMask: 255.255.255.0
  Multicast Specific:
    Multicast Capability: Yes
  GROUP             REFCNT
  -----             ------
  224.0.0.5         0000000001
  224.0.0.1         0000000001
  Link Statistics:
    BytesIn                          = 488
    Inbound Packets                  = 0
    Inbound Packets In Error         = 0
    Inbound Packets Discarded        = 0
    Inbound Packets With No Protocol = 0
    BytesOut                         = 1092
    Outbound Packets                 = 0
    Outbound Packets In Error        = 0
    Outbound Packets Discarded       = 0
```

The following example is a portion of the output of the Netstat DEvlinks/-d command with an IPv6 interface (OSAQDIO46) shown as ready.

```
IntfName: OSAQDIO46        IntfType: IPAQENET6 IntfStatus: Ready
    NetNum: n/a  QueSize: 0   Speed: 0000001000
   MacAddress: 0002559A3F65
    DupAddrDet: 1
    CfgRouter: Non                   ActRouter: Non
    CfgMtu: None                     ActMtu: 8992
  Multicast Specific:
    Multicast Capability: Yes
    RefCnt      Group
    ------      -----
    0000000001  ff02::1:ff03:1
    0000000002  ff02::1
  Interface Statistics:
    BytesIn                          = 592
    Inbound Packets                  = 0
    Inbound Packets In Error         = 0
    Inbound Packets Discarded        = 0
    Inbound Packets With No Protocol = 0
    BytesOut                         = 1008
    Outbound Packets                 = 0
    Outbound Packets In Error        = 0
    Outbound Packets Discarded       = 0
```

If the devices do not have a LnkStatus or IntfStatus of Ready, this must be resolved before continuing. There are several things that might cause the LnkStatus or IntfStatus to not be ready. For example, the device might not be defined to z/OS correctly, the device might not be defined in PROFILE.TCPIP correctly, and so on.

You can use the PING command to verify indirect routes exist to others hosts within the network .

```
ping 9.67.107.7
 CS V2R5: Pinging host 9.67.107.7
 Ping #1 response took 0.048 seconds.(48.012 milliseconds)
 READY
ping 2001:0db8:0:a1b:2:559a:3f65:3
 CS V2R5: Pinging host 2001:0db8:0:a1b:2:559a:3f65:3
```

```
 Ping #1 response took 0.051 seconds.(51.012 milliseconds)
 READY
```

Use the Traceroute command to verify that the correct route is being taken for each indirectly attached host:

```
tracerte 9.67.107.5
 CS V1R6: Traceroute to 9.67.107.5 (9.67.107.5)
 1 9.67.106.7 (9.67.106.7)  40 ms  7 ms  6 ms
 2 9.67.107.5 (9.67.107.5)  9 ms  8 ms  9 ms
 READY
```

The following example shows IPv6 for indirectly attached hosts:

```
tracerte 2001:0db8:0:a1c:2:36a4:b39a:7
 CS V1R6: Traceroute to 2001:0db8:0:a1c:2:36a4:b39a:7
 at IPv6 address: 2001:0db8:0:a1c:2:36a4:b39a:7
 1 fe80::1:2:3:4
 (fe80::1:2:3:4) 13 ms 25 ms 40 ms
 2 2001:0db8:0:a1c:2:36a4:b39a:7
 (2001:0db8:0:a1c:2:36a4:b39a:7) 29 ms 263 ms 196 ms
```

Chapter 7. Virtual IP Addressing

This topic contains information about the following subtopics:

- Terminology
- Introduction to VIPA
- Moving VIPA (Upon outage of TCP/IP)
- Static VIPAs, dynamic VIPAs (DVIPAs), and distributed dynamic VIPAs
- Using static VIPAs
- Using dynamic VIPAs (DVIPAs)
- Choosing which form of dynamic VIPA to use
- Configuring distributed DVIPAs — sysplex distributor
- Resolution of DVIPA conflicts
- IPv6 considerations
- Other considerations
- Example of configuring dynamic and distributed VIPAs
- Verifying the DVIPAs in a sysplex
- Verifying sysplex distributor workload
- DVIPAs and routing protocols

Note: This information applies to both IPv4 and IPv6, unless otherwise noted.

Terminology

Virtual IP Address (VIPA)
A VIPA is a generic term that refers to an internet address on a z/OS host that is not associated with a physical adapter. There are two types of VIPAs:

- A *Static VIPA* cannot be changed except through a VARY TCPIP,,OBEYFILE operator command.
- A *dynamic VIPA (DVIPA)* can move to other TCP/IP stack members in a sysplex or it can be activated by an application program, an instance of z/OS Container Extensions, or by a supplied utility. Dynamic VIPAs are used to implement sysplex distributor as described in "Considerations for VIPA" on page 55.

  A unique type of application instance dynamic VIPA, referred to as the IBM z/OS Container Extensions (zCX) DVIPA is created for each IBM z/OS Container Extensions address space. For more information about z/OS Container Extensions, see "IBM z/OS Container Extensions network overview" on page 110.

Distributed DVIPA
A distributed DVIPA, which is a special type of DVIPA, can distribute connections within a sysplex.

Dynamic routing
VIPAs are designed to interoperate with a dynamic routing daemon. Therefore, it is highly recommended that a routing daemon be used on a z/OS host that uses VIPAs.

Introduction to VIPA

Traditionally, an IP address is associated with each end of a physical link (or each point of access to a shared-medium LAN), and the IP addresses are unique across the entire visible network, which can be the Internet or a closed intranet. The majority of IP hosts have a single point of attachment to the network, but some hosts (particularly large server hosts) have more than one link into the network. A TCP/IP host with multiple points of attachment also has multiple IP addresses, one for each link.

© Copyright IBM Corp. 2000, 2022

Within the IP routing network, failure of any intermediate link or adapter disrupts user service only if there is not an alternate path through the routing network. Routers can route IP traffic around failures of intermediate links in such a way that the failures are not visible to the end applications or IP hosts. However, because an IP packet is routed based on ultimate destination IP address, if the adapter or link associated with the destination IP address fails, there is no way for the IP routing network to provide an alternate path to the stack and application. Endpoint (source or destination) IP adapters and links thus constitute single points of failure. While this might be acceptable for a client host, where only a single user will be cut off from service, a server IP link might serve hundreds or thousands of clients, all of whose services would be disrupted by a failure of the server link.

The virtual IP address (VIPA) removes the adapter as a single point of failure by providing an IP address that is associated with a stack without associating it with a specific physical network attachment. Because the virtual device exists only in software, it is always active and never experiences a physical failure. A VIPA has no single physical network attachment associated with it. Also, with the exception of a zCX DVIPA, the TCP/IP stack does not maintain interface counters for VIPA interfaces (VIRTUAL links).

To the routing network, a VIPA appears to be a host address indirectly attached to the z/OS. When a packet with a VIPA destination reaches the stack, the IP layer recognizes the address and passes it to the protocol layer in the stack.

The failure of the physical interface can be extended to the failure of the TCP/IP address space, the entire z/OS, or for planned outages. A VIPA just needs to move to a backup stack, and the routes to the VIPA need to be updated. Then clients can transparently connect to the backup stack. This process is known as VIPA takeover.

VIPA takeover improved with the introduction of dynamic virtual IP address (DVIPA) and distributed dynamic virtual IP address (distributed DVIPA). The DVIPA function improves VIPA takeover by allowing a system programmer to plan for system outages and provide for backup systems to take over without operator intervention or external automation. The distributed DVIPA function allows the connections for a single DVIPA to be serviced by applications on several stacks listed in the configuration statement (the distribution list). This adds the benefit of limiting the scope of an application or stack failure, while also providing enhanced work load balancing.

In general, z/OS configured with VIPA provides the following advantages:

- Automatic and transparent recovery from device and adapter failure.

  When a device (for example, a channel-attached router) or adapter (for example, an OSA-Express adapter) fails, if there is another device or link that provides the alternate paths to the destination:

  – IP will detect the failure, find an alternate path for each network, and route outbound traffic to hosts and routers on those networks via alternate paths.
  – Inbound and outbound traffic will not need to reestablish the active TCP connections that were using the failed device.
  – For connection requests originating at a z/OS TCP/IP stack, tolerance of device and adapter failures can be achieved by using source VIPA addressing.
- Recovery from z/OS TCP/IP stack failure (where an alternate z/OS TCP/IP stack has the necessary redundancy).

  Assuming that an alternate stack is installed to serve as a backup, the use of VIPAs enables the backup stack to activate the VIPA address of the failed stack.

  Connections on the failed primary stack will be disrupted but they can be reestablished on the backup using the same IP address as the destination. In addition, the temporarily reassigned VIPA address can be restored to the primary stack after the cause of failure has been removed.
- Limited scope of a stack or application failure.

  If a DVIPA is distributed among several stacks, the failure of only one stack affects only the subset of clients connected to that stack. If the distributing stack experiences the failure, a backup assumes control of the distribution and maintains all existing connections.
- Enhanced workload management through distribution of connection requests.

With a single DVIPA being serviced by multiple stacks, connection requests and associated workloads can be distributed across multiple z/OS images, according to Workload Manager (WLM) and Service Level Agreement policies (for example, QOS) or according to configured active connection distribution goals.

- Allows the non-disruptive movement of an application server to another stack so that workload can be drained from a system in preparation for a planned outage.

Moving a VIPA (for TCP/IP outage)

While a VIPA provides non-disruptive rerouting of IP data during failure of a physical interface, termination of the stack or the associated z/OS (including planned outages) will disrupt connections or UDP sessions to applications on the terminated stack. While failure of the TCP connection or UDP session will be visible to the clients, the duration of the outage is determined by how long the client application is unable to reconnect to an appropriate server application. Because it is common in large enterprises to have multiple instances of an application on different z/OS images, if the VIPA address can be moved to another stack that supports the application, the clients can reconnect and the perceived outage will be over.

An IP address associated with a particular physical device is unavailable until the owning stack is restarted; however, a VIPA is not associated with any particular physical interface. If termination of a stack is detected and a suitable application already is active on another stack, the VIPA can be moved. Connections on the terminated stack will be disrupted, but they can be reestablished on the backup stack using the original VIPA.

Movement of a static VIPA to a backup stack can be accomplished by using VARY TCPIP,,OBEYFILE commands on the backup. The data set specified on the command must contain an appropriate set of DEVICE, LINK, HOME, and optionally, BSDROUTINGPARMS statements for IPv4 static VIPAs or INTERFACE statements for IPv6 static VIPAs. If OMPROUTE is used as the routing daemon, an appropriate interface statement is needed in the OMPROUTE configuration file. If the TCP/IP configuration file with the statements defining the VIPA is created in advance, the transfer can be accomplished via automation. This procedure is documented in z/OS Communications Server: IP Configuration Reference. Movement of a DVIPA, on the other hand, can be accomplished by configuring a stack to back up a specific DVIPA that is defined on another stack. In this case, failure of the defining stack causes the DVIPA to move without operator intervention or extra automation. See “Planning for dynamic VIPA takeover” on page 369 for more information. Regardless of the type of VIPA to be moved, it is up to the system programmer or operator to ensure that the VIPA is moved to a backup stack that has the appropriate server applications.

In the absence of a failure, a VIPA is just like any other IP address, and routing for a VIPA is the same as for an IP address associated with a physical link.

Static VIPAs, dynamic VIPAs, distributed DVIPAs

z/OS TCP/IP stack supports two types of VIPAs: static and dynamic. Dynamic VIPAs (DVIPAs) can be used to distribute connections in a sysplex. This is referred to as a distributed DVIPA.

All three VIPAs can coexist on a given stack, but there are differences in how these VIPAs are configured and used.

Static VIPAs have the following characteristics:

- They can be activated during TCP/IP initialization or VARY TCPIP,,OBEYFILE command processing, and are configured using an appropriate set of DEVICE, LINK, HOME, and optionally, OMPROUTE configuration statements or BSDROUTINGPARMS statements for IPv4 Static VIPAs or INTERFACE statements for IPv6 Static VIPAs.
- Using the SOURCEVIPA configuration option, static VIPAs can be used as the source IP address for outbound datagrams for TCP, RAW, UDP (except routing protocols), and ICMP requests. For IPv6 static VIPAs to be used as source addresses, the SOURCEVIPA configuration option must be enabled and the VIPA interface must appear on the SOURCEVIPAINT keyword on some other INTERFACE statement.

This provides tolerance of device and adapter failures for connection requests originating at a z/OS TCP/IP stack.

- They can be specified as the source IP address for outbound TCP connection requests for all applications using this stack with TCPSTACKSOURCEVIPA.
- They can be specified as the source IP address for outbound TCP connection requests for specific jobs or specific destinations through the use of the SRCIP profile statement block.
- The number of static VIPAs on a stack is limited only by the range of host IP addresses that are available for that host.
- They can be moved to a backup stack after the original owning stack has failed, by using VARY TCPIP,,OBEYFILE command processing to configure the VIPA on the backup stack and updating the routers.

Dynamic VIPAs have the following characteristics:

- They can be configured to be moved dynamically from a failing stack to a backup stack within the same sysplex without operator intervention or external automation.
- They can be moved manually by deactivating or reactivating them with the VARY TCPIP,,SYSPLEX operator command.
- They can be dynamically activated by an application program.
- They can distribute connections within a sysplex.
- They can be specified on a TCPSTACKSOURCEVIPA statement. This allows a user to specify one VIPA to be used as the source IP address for outbound datagrams for TCP-only requests.
- They can be specified as the source IP address for outbound TCP connection requests for specific jobs or specific destinations by using the SRCIP profile statement block.
- Unlike static VIPAs, dynamic VIPAs:
  - Are limited to 1024 per stack.
  - Cannot be specified as the VIPA used by Enterprise Extender for connectivity purposes. (See “Configuring static VIPAs for Enterprise Extender” on page 368 for details.)

Distributed DVIPAs have the following characteristics:

- They have all the characteristics of DVIPAs, except that they cannot be dynamically activated by an application program.
- One stack defines a DVIPA and advertises its existence to the network. Stacks in the target distribution list activate the DVIPA and accept connection requests.
- Connection workload can be distributed across several stacks.

See “Configuring distributed DVIPAs - sysplex distributor” on page 382 for more detailed descriptions.

Guideline: OSA-Express devices have a limit on the number of IP addresses (both IPv4 and IPv6 addresses) that can be registered to the device. The limit is dependent on the microcode level of the OSA-Express device. This limit applies across all TCP/IP stacks that share the OSA-Express device. When defining a large number of VIPAs, take care not to exceed this limit. If the limit is exceeded, IP addresses beyond the limit will not be registered with the OSA-Express devices, and incoming packets with those IP addresses will not be routed to the correct stack unless that stack is designated as the primary router.

Using static VIPAs

The following topics describe how to configure static VIPAs, the special case of static VIPAs and Enterprise Extender, and how to implement static VIPA takeover.

Because a VIPA is associated with a z/OS TCP/IP stack and it is not associated with a specific physical network attachment, it can be moved to a stack on any image in the sysplex or even to any z/OS TCP/IP stack if the address fits into the network configuration.

Steps for configuring static VIPAs for a z/OS TCP/IP stack

A virtual IP address (VIPA) is an internet address on a z/OS host that is not associated with a physical adapter. A VIPA removes the adapter as a single point of failure by providing an IP address that is associated with a stack without associating it with a specific physical network attachment.

Procedure

Perform the following steps to configure a static VIPA address in one stack:

1. When configuring static VIPAs for the IPv4 network, add an INTERFACE statement of type VIRTUAL for each static VIPA to be defined. Alternatively, add DEVICE, LINK, HOME, and optionally BSDROUTINGPARMS statements for each static VIPA to be defined. When configuring static VIPAs for the IPv6 network, add INTERFACE statements of type VIRTUAL6 for each static VIPA to be defined.

   Notes:

   a. A VIPA link or VIPA interface cannot be coded on a static route in the BEGINROUTES statement.

   b. If you want to add a static VIPA in an IPv4 network with an address that already exists in the HOME list, you must first delete the existing address using the VARY TCPIP,,OBEYFILE command with the existing address omitted from the HOME list. Then use the VARY TCPIP,,OBEYFILE command with a new and complete HOME list.

2. For IPv4 networks, if tolerance of device and adapter failures is wanted for connection requests that originate at a z/OS TCP/IP stack, specify the SOURCEVIPA option on the IPCONFIG statement.

   Tips:

   - For the SOURCEVIPA option to work properly, the receiving nodes in the network must be configured to recognize the SOURCEVIPA addresses using the static or dynamic routing protocols. Otherwise, timeouts for the connection or request responses will occur as a result of the VIPA addresses being network unreachable.
   - If TCPSTACKSOURCEVIPA is specified on the IPCONFIG statement, it overrides SOURCEVIPA for outbound IPv4 TCP connections. If the SRCIP profile statement block is defined to establish one or more job-specific or destination-specific source IPv4 addresses, these IP addresses override TCPSTACKSOURCEVIPA or the VIPAs in the HOME list for IPCONFIG SOURCEVIPA, or both, for the specified job names and destinations. For information about the hierarchy of various ways that the source IP address of an outbound connection is determined, see “Source IP address selection” on page 244.

   For more information on configuring IPv4 SOURCEVIPA or TCPSTACKSOURCEVIPA addresses on the IPCONFIG statement, or SRCIP addresses in the SRCIP statement block, see z/OS Communications Server: IP Configuration Reference.

3. For IPv6 networks, if tolerance of device and adapter failures is wanted for connection requests that originate at a z/OS TCP/IP stack, specify the following option or keyword:

   - The SOURCEVIPA option on the IPCONFIG6 statement.
   - The SOURCEVIPAINT keyword with a VIPA interface name on the INTERFACE statements of the real (physical) interfaces, or on the DYNAMICXCF specification on the IPCONFIG6 statement.

   Tips:

   - For the SOURCEVIPA option to work properly, the receiving nodes in the network must be configured to recognize the SOURCEVIPAINT addresses using the static or dynamic routing protocols. Otherwise, timeouts for the connection or request responses will occur as a result of the VIPA addresses being network unreachable.
   - If the TCPSTACKSOURCEVIPA parameter is specified on the IPCONFIG6 statement, it overrides the SOURCEVIPA value for outbound IPv6 TCP connections. If the SRCIP profile statement block is defined to establish one or more job-specific or destination-specific source IPv6 addresses, these IP addresses override the TCPSTACKSOURCEVIPA value, the SOURCEVIPAINT value, or both, for the specified job names and destinations. For information about the hierarchy of various ways that the

source IP address of an outbound connection is determined, see “Source IP address selection” on page 244.

For more information on configuring IPv6 SOURCEVIPA or TCPSTACKSOURCEVIPA addresses on the IPCONFIG6 statement, or SRCIP addresses in the SRCIP statement block, see z/OS Communications Server: IP Configuration Reference.

4. For host name resolution of a VIPA address, configure the domain name servers to associate the host name with the VIPA.
5. Configure the routing daemon to advertise the presence of the VIPA.

   As described in prior steps, remember that the VIPA to be advertised can be determined by the SOURCEVIPA or TCPSTACKSOURCEVIPA parameters on the IPCONFIG and IPCONFIG6 statements, or by the SRCIP statement. For more information, see z/OS Communications Server: IP Configuration Reference.

Results

Figure 50 on page 366 illustrates a simple configuration showing multiple network attachments using a single static VIPA address. Because any other network interface can be used with static VIPA's, see “Setting up physical characteristics in PROFILE.TCPIP” on page 250 for descriptions of other network interfaces. The simple configuration will be used as the TCPCS6 system throughout this information.

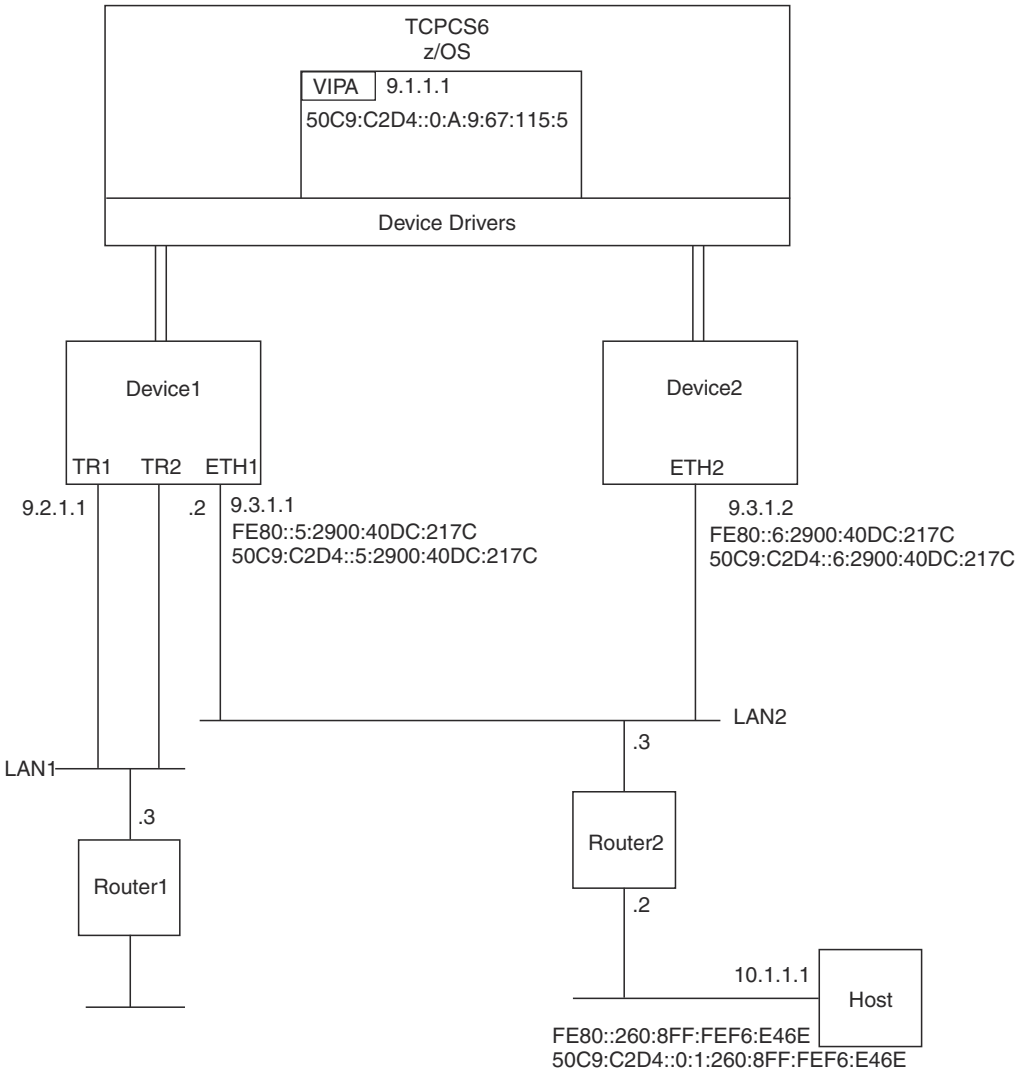


Figure 50. Static VIPA configuration

Steps for converting from IPv4 VIRTUAL DEVICE, LINK, and HOME definitions to the IPv4 VIRTUAL INTERFACE statement

You can use the INTERFACE statement in the TCP/IP profile to configure IPv4 definitions for static VIPAs rather than using the DEVICE, LINK, and HOME statements. Using the INTERFACE statement simplifies stack configuration for VIRTUAL interfaces.

Before you begin

If a static VIPA is being used as a source VIPA for any IPv4 interfaces that are configured using DEVICE and LINK statements, then you cannot convert that static VIPA to an IPv4 INTERFACE statement.

Procedure

Perform the following steps to convert your TCP/IP profile so that it uses the INTERFACE statement to configure IPv4 definitions for static VIPAs.

1. Convert the IPv4 VIRTUAL DEVICE, LINK, and HOME statements to an IPv4 VIRTUAL INTERFACE statement.

   The values used in the following substeps are based on the sample profile shown in Figure 51 on page 367.

   a. Copy the LINK name (VIPA1 in the example) and specify this name as the interface name on the INTERFACE statement.

      Tip: When you use the original LINK name as the new INTERFACE name, you do not have to change OMPROUTE definitions or the PRIMARYINTERFACE statement.

   b. Copy the VIRTUAL parameter from the LINK statement and specify this parameter after the DEFINE parameter of the INTERFACE statement.

   c. Copy the HOME list entry IP address (201.16.1.1 in the example) and specify this address on the IPADDR parameter of the INTERFACE statement.

   d. Remove the HOME list entry.

   Example of the profile before conversion:

```
DEVICE VIPA    VIRTUAL   0
LINK      VIPA1  VIRTUAL   0   VIPA
;
HOME
 201.16.1.1     VIPA1
```

   Example of the profile after conversion:

```
INTERFACE VIPA1
 DEFINE VIRTUAL
 IPADDR 201.16.1.1
```

   Figure 51. Examples of the profile before and after conversion

   Tip: Optionally, you can dump the TCP/IP address space and use the CONVERT parameter on the TCPIPCS PROFILE subcommand to display the configuration information at the time of the dump. The resulting output reflects your IPAQENET, IPAQIDIO, and VIRTUAL DEVICE, LINK, and HOME definitions in INTERFACE statement format, which might be helpful in converting your profile to use INTERFACE statements. Review the output before you implement any changes. For more information about using the CONVERT parameter on the TCPIPCS PROFILE subcommand, see z/OS Communications Server: IP Diagnosis Guide.

   For more information about the IPv4 VIRTUAL INTERFACE statement, see z/OS Communications Server: IP Configuration Reference.

2. Remove any BSDROUTINGPARMS entries for the interface.

Results

For information about the TCP/IP profile (PROFILE.TCPIP) and configuration statements, see z/OS Communications Server: IP Configuration Reference.

Configuring static VIPAs for Enterprise Extender

Defining at least one static VIPA is required by VTAM to access the IP network. Because VTAM does not move within a sysplex, a dynamic VIPA cannot be used. Enterprise Extender supports the use of multiple static VIPA addresses, and a VIPA address is chosen as follows:

1. VTAM uses the IPADDR value specified on the Enterprise Extender XCA major node GROUP definition statement, or the VIPA address obtained when name-to-address resolution is performed on the HOSTNAME value specified on the Enterprise Extender XCA major node GROUP definition statement.
2. If the GROUP definition statement specifies neither IPADDR nor HOSTNAME, VTAM uses the static VIPA address specified on the VTAM IPADDR start option, or the static VIPA address returned by the resolver when name-to-address resolution is performed on the value of the VTAM HOSTNAME start option.
3. If neither the IPADDR nor the HOSTNAME start option is specified, VTAM uses the first IPv4 static VIPA in the HOME list.

If remote APPN nodes use a host name and not a host address to define the destination of an Enterprise Extender connection, the domain name server must return the VIPA address used by VTAM for the host name. Alternatively, if the Enterprise Extender endpoints are behind a firewall or for another reason require network address translation, the domain name server should return a network address translation (NAT) address. This NAT address should in turn map, on the destination side, to the static VIPA address of the intended target VTAM host.

Rule: If you plan on using IPv4 protocols for your Enterprise Extender communication, you must define DEVICE, LINK, and optionally HOME statements for IUTSAMEH, or specify IPCONFIG DYNAMICXCF in the appropriate TCPIP profile data set. If you plan on using IPv6 protocols for your Enterprise Extender communication, you must define an INTERFACE statement for IUTSAMEH or specify IPCONFIG6 DYNAMICXCF in the appropriate TCPIP profile data set.

For more information about using Enterprise Extender, see z/OS Communications Server: SNA Network Implementation Guide.

Considerations when using static VIPAs with IPv6

When static VIPAs are configured for use with IPv6, it is recommended that the prefixes of the IPv6 VIPA addresses be different than the prefixes used for addresses assigned to real interfaces. This reduces the likelihood of address collisions between the manually configured VIPA addresses and the autoconfigured addresses of the real interfaces.

To allow other hosts that share links with the z/OS TCP/IP stack to access the IPv6 VIPA addresses, without the need for manual route configuration, a router on each of the links should include the VIPA prefix in its router advertisements. The router advertisements should define the prefix as being on-link and should indicate that the prefix should not be used for autoconfiguration.

Planning for static VIPA takeover and takeback

Because a VIPA is associated with a z/OS TCP/IP stack and is not associated with a specific physical network attachment, it can be moved to a stack on any image in the sysplex or even to any z/OS TCP/IP stack as long as the address fits into the network configuration. Moving a static VIPA can be done manually by an operator or by customer-programmed automation. Movement of the static VIPA allows other hosts that were connected to the primary stack to reestablish connections with a backup TCP/IP stack using the same VIPA. After the primary TCP/IP stack has been restored, the reassigned VIPA address can be moved back.

Consider the following items when backing up and restoring a z/OS TCP/IP stack:

- All connections on the failing host will be disrupted.
- The client can use any ephemeral port number when reestablishing the connection to the backup server.
- Having a different port number for the backup and primary server is not recommended. For example, if the primary FTP used port 21 and the backup FTP used port 1021, when backing up and restoring a z/OS TCP/IP stack, the client would have to know whether to use port 21 or 1021.

If you are deploying both static VIPA and OSA-Express QDIO, see "VIPAs, OSA-Express QDIO, and Spanning Tree Protocol" on page 415 to determine whether or not network bridge or switch configuration settings might potentially impact the environment for VIPA takeover and takeback.

Using dynamic VIPAs

Dynamic VIPA (DVIPA) support allows:

- Dynamic movement of a VIPA from a failing TCP/IP stack to a backup stack
- Dynamic allocation of a VIPA by an application program
- Manual movement of a VIPA by deactivating or reactivating it with the VARY TCPIP,,SYSPLEX command

Dynamic VIPAs (DVIPAs) are IP addresses like all other IP addresses associated with a TCP/IP, and they appear as though they had been defined at the end of the HOME list.

Dynamic VIPAs can be either IPv4 or IPv6, and both can be configured within the same VIPADYNAMIC block. A single statement (for example, VIPADEFINE or VIPABACKUP) must contain either IPv4 addresses or IPv6 addresses, but not both. However, statements that contain IPv4 addresses can be intermixed with statements that contain IPv6 addresses within the same VIPADYNAMIC block in any manner wanted.

Configuring DVIPA support

Unlike static VIPAs, DVIPAs are not configured by using DEVICE, LINK, and HOME statements (for IPv4) or by using INTERFACE statements (for IPv6). The configuration statements for the DVIPA support are contained within the VIPADYNAMIC block and consist of the following statements:

- VIPADEFINE and VIPABACKUP statements that are used to configure DVIPAs to be dynamically moved from a failing TCP/IP to a backup TCP/IP
- VIPARANGE used to specify a range of IP addresses which can be dynamically activated as a VIPA by:
  - an application program
  - a z/OS Container Extensions address space (using the VIPARANGE ZCX keyword)
- VIPADELETE used to delete existing DVIPAs
- VIPADISTRIBUTE used to configure a DVIPA as a distributed DVIPA and designate the target stacks

The following topics discuss how these statements are used to provide the DVIPA support. For syntax details, see z/OS Communications Server: IP Configuration Reference. For more information, see *IBM z/OS V1R13 Communications Server TCP/IP Implementation, Volume 3: High Availability, Scalability, and Performance*, SG24–7998.

When dynamically created DVIPAs are deleted, any applications that are bound to those DVIPAs (VIPARANGE or MODDVIPA) receive an asynchronous error, EUNATCH (3448) - the protocol required to support the address family is unavailable.

Planning for dynamic VIPA takeover

Movement by network management automation or operator intervention is not always desirable. Operator intervention takes time and is subject to errors. Automation requires correct detection of the failure and is also prone to error if the failure does not produce the exact console messages anticipated.

The dynamic VIPA takeover function addresses this problem. It is important to understand that dynamic VIPA takeover does not introduce functions that could not be accomplished by operator action or

automation. It just removes the dependency on human detection of the error or customer programming for automation. Dynamic VIPA takeover is completely accomplished by the TCP/IP stacks.

DVIPA takeover is possible when a DVIPA is configured as active (using VIPADEFINE) on one stack and as backup (using VIPABACKUP) on another stack within the sysplex. When the stack on which the DVIPA is active terminates or leaves the sysplex group, the backup stack automatically activates the DVIPA and notifies the routing daemon. For information on what causes a stack to leave the sysplex group, see "Sysplex problem detection and recovery" on page 452.

For DVIPA takeover to be useful, the applications that service the DVIPA addresses must be available on the backup stacks. In the absence of the application, the DVIPA will be active, but client connections to the application will still fail. If OMPROUTE is used, it is recommended that GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN be configured. This causes DVIPA takeover to be delayed until OMPROUTE is active and able to advertise DVIPAs on the takeover stack. For more information on how DELAYJOIN works, see "Sysplex problem detection and recovery" on page 452.

To preserve connections during DVIPA takeover, the TCP/IP stacks require XCF links. The DYNAMICXCF option must be coded in the TCP/IP profile of both stacks.

A determination of how the workload will be distributed among the backup stacks when the primary stack fails should be made. It is possible to designate a single stack as a backup and move all the workload to it, or the workload can be distributed among several stacks. In the first case, you need to configure only one DVIPA with a VIPADEFINE statement on the primary stack, and only one VIPABACKUP statement is required on the backup stack. For the second option, you must configure multiple DVIPAs with a VIPADEFINE statement on the primary stack.

After determining the workload distribution, each of the secondary stacks will require a VIPABACKUP statement for the DVIPA it will be supporting.

The following example shows how to implement a single stack backup for multiple applications.

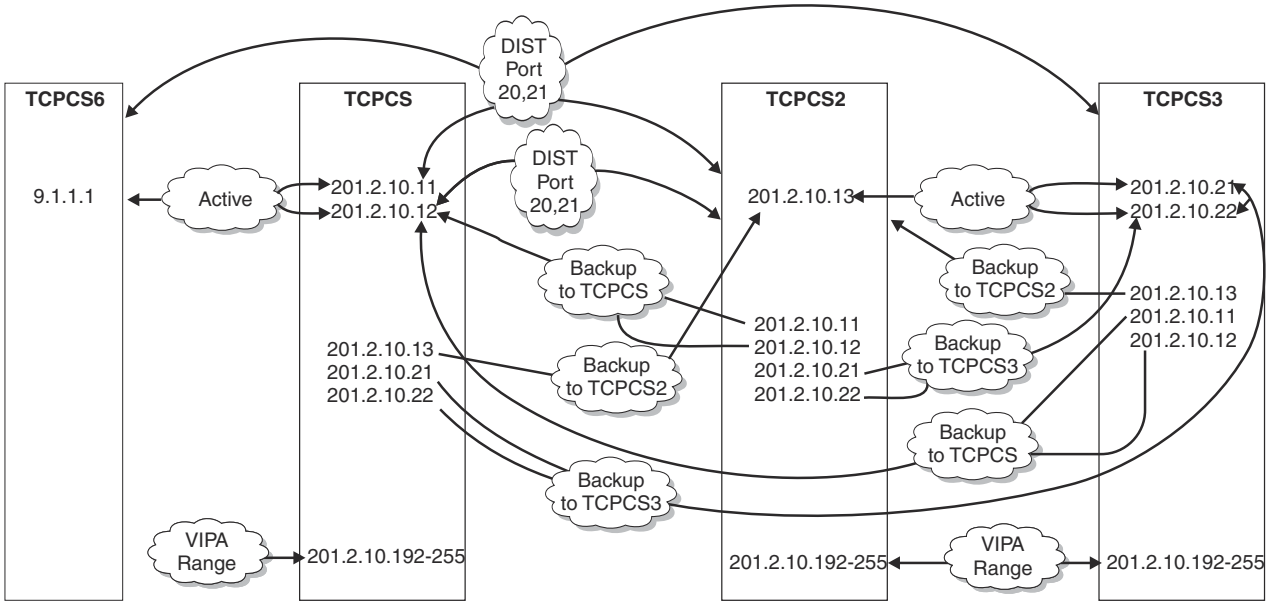


Figure 52. Sample DVIPA addressing in a sysplex environment

Stack TCPCS:

```
Uses VIPADEFINE to define 201.2.10.11
Has a Web server running that binds to INADDR_ANY.
  Web client programs use 201.2.10.11 as their destination address.
Has an FTP server running that binds to INADDR_ANY.
  FTP client programs use 201.2.10.11 as their destination address.
```

Stack TCPCS3:

```
Uses VIPABACKUP to define 201.2.10.11 as backup for stack TCPCS.
Has a Web server running that binds to INADDR_ANY.
Has an FTP server running that binds to INADDR_ANY.
```

In the preceding scenario, when stack TCPCS goes down, stack TCPCS3 receives all new connection requests for both the web and FTP servers. FTP and web client programs continue to use 201.2.10.11 as their destination address, but they now connect to stack TCPCS3.

The following example shows how to implement a multiple stack backup for multiple applications.

Stack TCPCS:

```
Uses VIPADEFINE to define 201.2.10.11 and 201.2.10.12
Has a Web server running that binds to INADDR_ANY.
  Web client programs use 201.2.10.11 as their destination address.
Has an FTP server running that binds to INADDR_ANY.
  FTP client programs use 201.2.10.12 as their destination address.
```

Stack TCPCS2:

```
Uses VIPABACKUP to define 201.2.10.11 as backup for stack TCPCS.
Has a Web server running that binds to INADDR_ANY.
```

Stack TCPCS3:

```
Uses VIPABACKUP to define 201.2.10.12 as backup for stack TCPCS.
Has an FTP server running that binds to INADDR_ANY.
```

In the preceding scenario, when stack TCPCS goes down, new connections for the web server at 201.2.10.11 will connect with stack TCPCS2, and new connections for the FTP server at 201.2.10.12 will connect with stack TCPCS3.

Manually initiating takeover for an individual dynamic VIPA

There might be times when you want a takeover to occur for an individual DVIPA for temporary, operational purposes, such as dealing with temporary shifts in available capacity. An operator command, VARY TCPIP,,SYSPLEX,DEACTIVATE, is provided to deactivate an active VIPADEFINE DVIPA so that it appears to have been deleted from the stack. This enables an already-configured backup stack to takeover the DVIPA. This command also saves the original configuration definition.

When you want to move the DVIPA back to the original stack, you can issue another operator command, VARY TCPIP,,SYSPLEX,REACTIVATE, on the original stack to cause the DVIPA to be redefined with its saved configuration. This causes a VIPADEFINE DVIPA to be activated again on the original stack, and causes the backup stack to relinquish ownership of the DVIPA and return to being a backup stack.

The following example shows how to deactivate a DVIPA:

```
VARY TCPIP,,SYSPLEX,DEACTIVATE,DVIPA=203.1.1.99
```

The following example shows how to reactivate a DVIPA that has been deactivated:

```
VARY TCPIP,,SYSPLEX,REACTIVATE,DVIPA=203.1.1.99
```

Deactivating a DVIPA can be done only for DVIPAs that have been configured on the stack with VIPADEFINE or VIPABACKUP. You cannot deactivate a VIPARANGE DVIPA created by BIND, the SIOCSVIPA or SIOCSVIPA6 ioctl command, or the MODDVIPA utility. When you deactivate a DVIPA, if there are any existing connections to the DVIPA on this stack and there is another stack able to maintain the connections, the DVIPA is kept in QUIESCING status until the last connection terminates, and then the DVIPA is deactivated.

You can also use these operator commands to deactivate and reactivate a backup DVIPA. If you deactivate a DVIPA that is in backup status, it makes that stack ineligible to takeover the DVIPA. Reactivating a VIPABACKUP DVIPA makes the stack eligible again to takeover the DVIPA.

While a DVIPA is deactivated it still appears in the Netstat VIPADCFG/-F report, where it is identified as deactivated, but it does not appear in any other Netstat reports unless the DVIPA is in QUIESCING status or this stack is a target for that DVIPA from some other distributing stack.

Different application uses of IP addresses and DVIPAs

Not all IP-based server applications relate to IP addresses in the same way. Automated movement of DVIPAs, and the planning for dynamic VIPA takeover, must take this difference into account.

Some IPv4 or IPv6 applications, such as FTP or the TN3270E Telnet server (Telnet) will accept client requests on any IP address by binding to INADDR_ANY or the IPv6 unspecified address (in6addr_any). The distinguishing feature of such an application is the function it provides, such as the particular set of SNA applications for Telnet. If the function is replicated across multiple z/OS images in the sysplex, as is often the case for distributed workload, the DVIPA must merely be moved to a stack supporting the application. This scenario is called the multiple application-instance scenario. For the multiple application-instance scenario, the stacks in the sysplex do all the work of activating a DVIPA in the event of a failure.

For other types of applications, each application instance must have a unique IP address. This scenario is called the unique application-instance scenario and uses DVIPAs that are activated with an ioctl or a bind().

To maintain the relationship between an application instance and its DVIPA, the application must indicate to the stack that the DVIPA needs to be activated. This occurs in the following cases:

- When the application instance issues a bind() function call and specifies an IP address that is not active on the stack. The stack will activate the address as a DVIPA, provided it meets certain criteria. When the application instance closes the socket, the DVIPA is deleted.
- Some applications cannot be configured to issue bind() for a specific IP address, but are unique application-instance scenario applications. For such applications, a utility is provided (MODDVIPA), which issues SIOCSVIPA or SIOCSVIPA6 ioctl() to activate or deactivate the DVIPA. This utility can be included in a JCL procedure or OMVS script to activate the DVIPA before initiating the application. As long as the same JCL package or script is used to restart the application instance on another node in the event of a failure, the same DVIPA will be activated on the new stack. For information about the authorization required to execute the MODDVIPA utility, see “Using the MODDVIPA utility” on page 376.
- An alternative for unique application-instance scenario server applications that cannot themselves bind to a unique IP address is to use the BIND parameter on the PORT reservation statement. It is always a good practice to reserve a port for the listening socket of a server application. If the BIND parameter and an IP address are specified on the PORT reservation statement for a server application, and the application binds a socket to that port and either the IPv4 INADDR_ANY address or the IPv6 unspecified address (in6addr_any), z/OS TCP/IP will convert the bind to be specific to the IP address specified on the PORT reservation statement. From that point on, it will appear as though the application itself had issued the bind() specifying the designated IP address, including having the IP address deleted when the server application closes the socket.

Configuring dynamic VIPAs

To allow continued and unchanged operation of static VIPAs in z/OS TCP/IP, DVIPAs are defined with configuration statements in the PROFILE.TCPIP data set. An overview of the relevant configuration statements is provided in the following topics, and also see “Verifying the DVIPAs in a sysplex” on page 420 for a description of the configuration statements. For an example of the VIPADYNAMIC configuration statement and display commands for dynamic VIPAs, see z/OS Communications Server: IP Configuration Reference.

Configuring the multiple application-instance scenario

For the multiple application-instance scenario, each instance is assigned a unique DVIPA. The VIPADEFINE keyword of the VIPADYNAMIC configuration statement is used to create the DVIPA on the stack where the DVIPA is normally expected to be active. When the VIPADEFINE statement is processed in a TCP/IP profile, corresponding DEVICE, LINK, HOME, and BSDROUTINGPARMS statements are generated automatically for IPv4 addresses. For IPv6 addresses, an INTERFACE statement is automatically generated. Routing daemons are automatically informed.

Additional configuration is required on other stacks in the sysplex to indicate which stack should take over the DVIPA in the event of a failure. The VIPADYNAMIC statement has a VIPABACKUP keyword for this purpose. A VIPABACKUP configures the DVIPA but does not activate it until it is necessary. Because more than one TCP/IP can back up a single DVIPA, a rank parameter on the VIPABACKUP statement determines the order in which several stacks will assume responsibility for a DVIPA.

The stacks in the sysplex exchange information on all VIPADEFINEs and VIPABACKUPs defined in the sysplex, so that all are aware of which stack should take over a particular DVIPA. The list of backup stacks for a specific DVIPA can be different from the list of backup stacks for all other DVIPAs.

In the multiple application-instance scenario, instances of the application in question are activated among sysplex nodes according to some plan, presumably related to balancing workload across available capacity. This activation is done independently of VIPA takeover. Configure the associated DVIPAs as follows:

1. For each instance of a particular application to be supported via DVIPA, add a VIPADEFINE statement to the TCP/IP profile for the TCP/IP associated with the application instance.
2. For each of the dynamic VIPAs in step 1, determine which application instance or instances should take over the workload (considering probable capacity and any other application-related considerations). If more than one TCP/IP is to provide backup for a DVIPA, determine the order in which the selected TCP/IPs should be designated as backup. Add a VIPABACKUP statement to each TCP/IP that is to provide backup for the DVIPA, with appropriate rank values to determine the order. Do this for each of the DVIPAs in step 1.
3. Perform steps 1 and 2 for each other application to be supported by DVIPAs.

   Note: It is possible to share a dynamic VIPA among several different applications, but in doing so, ensure that instances of all such applications will exist together on any TCP/IP to which the DVIPA may be moved in case of a failure.

After these steps are complete, start the affected TCP/IPs (or modify their configuration using the VARY TCPIP,,OBEYFILE command), if applicable, configure DNS for the application names, and start the application instances. From that point on, the TCP/IPs in the sysplex will collaborate to ensure that each dynamic VIPA is kept active somewhere within the sysplex as long as there is at least one functioning TCP/IP which has been designated as backup for the dynamic VIPA.

Note: The limit of 1024 DVIPAs on a single TCPIP applies to the DVIPAs that are defined by the VIPADEFINE or VIPABACKUP configuration statements or by a VIPADISTRIBUTE statement on another stack.

Configuring the unique application-instance scenario

The unique application-instance scenario ties a DVIPA to a specific instance of an application. To isolate errors in configuring applications, TCP/IP needs a mechanism to identify permissible DVIPAs. This is provided with one or more VIPARANGE statements. The VIPARANGE statement identifies a range of IP addresses which can be activated as DVIPAs by an application instance. Each TCP/IP stack can have up to 4096 VIPARANGE definition statements for both IPv4 and IPv6. The VIPARANGE statement consists of an IPv4 address and subnet mask, or an interface name and an IPv6 address and prefix length, and defines a subnet for DVIPAs. More than one VIPARANGE statement with different ranges can be defined on a TCP/IP. VIPARANGE does not itself cause any DVIPAs to be activated. Rather, DVIPAs are activated either by an application issuing a bind() for a specific IP address, by use of the SIOCSVIPA or SIOCSVIPA6 ioctl() command issued by an authorized application, or by the MODDVIPA utility.

When an application issues bind() for a specific IP address or an address was selected using the BIND keyword on the PORT statement, the receiving stack checks it against addresses in the HOME list. If the IP address has already been activated on this stack (whether for a physical device, a static VIPA, or a dynamic VIPA), the bind() execution is successful. If the IP address is not active on this TCP/IP, the current VIPARANGE statements are checked to see if the IP address falls within one of them. If an appropriate VIPARANGE is found, the IP address is activated as a DVIPA and the operation succeeds. If no appropriate VIPARANGE is found, or if the IP address is active elsewhere in the sysplex other than by a NONDISRUPTIVE bind(), the request fails and bind() returns EADDRNOTAVAIL.

When an authorized application issues the SIOCSVIPA or SIOCSVIPA6 ioctl() command to create a DVIPA, or when the MODDVIPA utility is executed in JCL or an OMVS script to activate a DVIPA on behalf of an application instance, the current VIPARANGE statements are checked to see whether the IP address falls within one of them. If an appropriate VIPARANGE is found, and the IP address is not currently active on this TCP/IP as a VIPADEFINE/VIPABACKUP dynamic VIPA, then the IP address is activated as a DVIPA. However if no appropriate VIPARANGE is found on this TCP/IP, or if the IP address is currently defined on this TCP/IP or configured elsewhere in the sysplex as a VIPADEFINE/VIPABACKUP dynamic VIPA, then the request fails with errno and errnojr set to indicate the reason for the failure and the utility ends with a nonzero condition code. See “Dynamic VIPA creation results” on page 405 for more information.

Tip: If the requested IP address has been activated as a dynamic VIPA by a bind(), SIOCSVIPA ioctl, or SIOCSVIPA6 ioctl elsewhere in the sysplex, the result depends on how the stacks were configured. See “Dynamic VIPA creation results” on page 405 for more information.

Application-instance DVIPAs provide high availability features for applications and workloads that exploit them. The creation and movement of these DVIPAs is quite flexible, as it can be triggered without modification to the TCP/IP profile and without issuing TCP/IP operator commands. While this flexibility is useful, it can also present problems for the network administrator in determining when and how a DVIPA is activated on a TCP/IP stack.

To help alleviate these problems, TCP/IP provides auditing facilities to track the state changes of application-instance DVIPAs. When an application-instance DVIPA is created or deleted dynamically through an explicit bind, by executing the MODDVIPA utility or by invoking the SIOCSVIPA or SIOCSVIPA6 ioctl command, messages are issued (EZD1204I, EZD1205I, EZD1206I, or EZD1207I) to the job log of the TCP/IP started procedure and to the system log. These messages include the method (bind, ioctl, or MODDVIPA utility) used to create or delete the DVIPAs, and the job name at the time the bind, ioctl, or MODDVIPA utility was invoked. For more information about EZD1204I, EZD1205I, EZD1206I, and EZD1207I, see z/OS Communications Server: IP Messages Volume 2 (EZB, EZD).

The Netstat VIPADyn/-v report is enhanced for application-instance DVIPAs to include the job name and the time stamp that this DVIPA was activated on the local TCP/IP stack (either because it is the owner of the DVIPA or is a target for this DVIPA). For more details about Netstat VIPADyn/-v, see z/OS Communications Server: IP System Administrator's Commands.

In the unique application-instance scenario, each application instance is assigned a unique IP address as its DVIPA. Before defining individual DVIPAs, one or more blocks of IP addresses must be defined for these DVIPAs, and the individual DVIPAs must be defined from within the blocks. Each block should be represented as a subnet, so that a VIPARANGE statement can be defined for it.

Follow these steps when setting up any unique application instances:

1. For each application instance, assign a DVIPA from one of the blocks of IP addresses for this purpose. Do not assign an IP address which is also assigned to another application instance, or which is defined by VIPADEFINE for the multiple application-instance scenario. Configure the application to use this DVIPA [if it issues bind() for a particular IP address], or add a BIND parameter to the PORT reservation statement for the port of the application's listening socket to cause the listening socket to be bound to this DVIPA. Alternatively, add the MODDVIPA utility to the JCL or OMVS script and configure the MODDVIPA utility to activate the DVIPA before starting the application. When the application ends, use the MODDVIPA utility to delete the DVIPA.
2. For each application instance, determine on which stack the application instance will normally be executed and to which stacks the application instance could be moved in case of failure of the normal stack or the application itself. For each such stack, add a valid VIPARANGE statement to the profile.

Note: The dynamic VIPA must be within the VIPARANGE subnet. The broadcast address and the net prefix cannot be used.

3. Perform steps 1 and 2 until all application instances have been allocated a unique DVIPA.

The application restart strategy should ensure that the worst-case failure scenario does not attempt to activate more than 4096 DVIPAs on a single stack. If such an attempt is made, activation of the 4097th DVIPA fails, with possible resulting loss of connectivity from clients to the server application.

Note: The limit of 4096 DVIPAs on a single TCP/IP applies to all DVIPAs, whether the DVIPAs are defined by the VIPADEFINE or VIPABACKUP configuration statements, by a VIPADISTRIBUTE statement on another stack, by a bind() call, or by executing the MODDVIPA utility.

Defining a single block makes the definition process easier, but also provides less individual control. Alternatively, because the smallest subnet consists of four IP addresses, defining a unique subnet for each DVIPA in this scenario wastes three other IP addresses that could have been used for DVIPAs.

Use of the SIOCSVIPA or SIOCSVIPA6 ioctl command

An ioctl command, SIOCSVIPA or SIOCSVIPA6, allows an application to create or delete a dynamic VIPA on the stack where the application is running. The application can issue the IOCTL call by using the run time XL C/C++ or by using the UNIX System Services Callable Services BPX1IOC API. The application that is issuing the SIOCSVIPA or SIOCSVIPA6 ioctl command must be APF-authorized. If there is no security product, or if the security product indicates that no profile for MODDVIPA is defined, the application must be running under a user ID with superuser authority. If the profile for the MODDVIPA program is defined, the user ID must be permitted READ access to the profile even if the ID has superuser authority. Some security products indicate that there is no access available when no profile is defined. If you are not using RACF, you might need to define the profile to use these ioctl commands. For more information, see "Defining a security profile for SIOCSVIPA, SIOCSVIPA6, and MODDVIPA" on page 377.

To create a new dynamic VIPA, the requested IP address must be within a subnet that is previously specified by a VIPARANGE configuration statement in the PROFILE.TCPIP data set for this stack. The SIOCSVIPA or SIOCSVIPA6 ioctl command can be used to delete any existing dynamic VIPA on the stack, except for distributed DVIPAs. A dynamic VIPA can be created by using the DVR_DEFINE option or the DVR_DEFINE_AFFINITY option. The DVR_DEFINE_AFFINITY option creates the DVIPA with affinity to the address space that created it. A connection request for a DVIPA that is created with affinity is sent to a TCP listener if the application instance that created the DVIPA issues the bind() call. If no matching listener is found, a TCP listener is selected by using the normal shareport load balancing.

The following example shows how to set up the SIOCSVIPA ioctl command for applications that are to use IPv4 addresses.

```
#include "ezbzdvpc.h"               /* header that contains
                                      the structure for
                                      SIOCSVIPA ioctl
                                       and needed constants*/
struct dvreq dv;                    /* the structure passed
                                         on the ioctl command*/
dv.dvr_version = DVR_VER1;           /*version               */
dv.dvr_length = sizeof(struct dvreq);  /* structure length     */
dv.dvr_option = DVR_DEFINE;           /* to define a new
                                         dynamic VIPA.  Use
                                         DVR_DELETE to delete
                                         a dynamic VIPA        */
dv.dvr_addr.s_addr = inet_addr(my_ipaddr);  /* where my_ipaddr is
                                               a character string
                                               in standard
                                               dotted-decimal
                                               notation */
```

The ioctl command is then invoked as follows:

```
rc = ioctl(s, SIOCSVIPA, &dv);
```

The following example shows how to set up the SIOCSVIPA6 ioctl command with the input parameter list that supports IPv6 addresses.

```
#include "ezbzdvpc.h"           /* header that contains
                                   the structure for
                                   SIOCSVIPA6 ioctl
                                   and needed constants */
 struct dvreq6 dv6;              /* the structure passed on
                                   the ioctl command */
 dv6.dvr6_version = DVR_VER2;   /* version */
 dv6.dvr6_length = sizeof(struct dvreq6); /* structure length */
 dv6.dvr6_option = DVR_DEFINE;  /* to define a new dynamic
                                   VIPA.  Use DVR_DELETE to
                                   delete a dynamic VIPA. */
inet_pton(AF_INET6, my_ipv6addr, dv6.dvr6_addr6.s6_addr);
                                /* where my_ipv6addr
                                   is a character string in
                                   standard IPv6 address
                                   notation, representing a
                                   fully qualified IPv6
                                   address */
```

The ioctl command is then invoked as follows:

```
rc = ioctl(s, SIOCSVIPA6, &dv6);
```

The SIOCSVIPA or SIOCSVIPA6 ioctl command sets nonzero errno and errnojr values to indicate error conditions. See z/OS Communications Server: IP and SNA Codes for a description of the errnojr values returned.

Using the MODDVIPA utility

You can use the MODDVIPA utility to activate or delete a dynamic VIPA. The utility can be initiated from JCL or an OMVS script. MODDVIPA must be loaded from an APF-authorized library and be run under a user ID that has an OMVS segment. If there is no security product, or if the profile for MODDVIPA has not been defined, the application must be running under a user ID with superuser authority. If the profile for the MODDVIPA program has been defined, the user ID must be permitted READ access to the profile even if the ID has superuser authority. For more information, see “Defining a security profile for SIOCSVIPA, SIOCSVIPA6, and MODDVIPA” on page 377.

Restriction: The MODDVIPA utility cannot be used to create or delete a dynamic VIPA from a VIPARANGE statement for zCX.

Input parameters

The input parameters for the utility are:

-p <*tcpipname*>
Specifies the TCP/IP which is to create or delete a DVIPA.

-c <*IPaddress*>
Specifies that the IP address is to be created.

-a <IPaddress>
Specifies that the IP address is to be created with affinity. A connection request for a DVIPA that is created with affinity is sent to a TCP Listener if the application instance that created the DVIPA issues the bind() call. If no matching listener is found, a TCP Listener is selected by using normal shareport load balancing.

-d <IPaddress>
Specifies that the IP address is to be deleted.

Notes:

1. The input parameters `-p`, `-c`, `-a`, and `-d` must be entered in lowercase.
2. <*tcpipname*> must be entered in uppercase.

3. For IPv4 addresses, <*IPaddress*> is dotted decimal notation. For IPv6 addresses, <*IPaddress*> is standard colon-hexadecimal notation for specifying IPv6 addresses.
4. To create a DVIPA, the specified IP address must be within a subnet that is previously specified by a VIPARANGE configuration statement in the PROFILE.TCPIP data set for the specified TCP/IP.

Output

The MODDVIPA utility sets the following exit (completion) codes for create (`-c`):

0
Success: The DVIPA was activated.

4
Warning: The requested DVIPA is within the VIPARANGE subnet but it is already active on this stack; another application created this DVIPA. Your application can use the DVIPA, but the application that created the DVIPA can delete it at any time.

8
Error: The IP address was not defined as a DVIPA on this TCP/IP.

12
An error was found in the input parameters

The MODDVIPA utility sets the following exit (completion) codes for delete (`-d`):

0
Success: The dynamic VIPA was deleted.

8
Error: The requested DVIPA was not deleted.

12
An error was found in the input parameters

Notes:

1. When an error is detected, the ernno text and errnojr value are printed to stderr.
2. If the IP address requested for the DVIPA is not within a VIPARANGE configured on this stack, completion code 8 is returned even if the IP address is currently active on this stack

Examples

Within JCL:

```
//TCPDVP EXEC  PGM=MODDVIPA,REGION=0K,TIME=1440,                        X
//             PARM='POSIX(ON) ALL31(ON)/-p TCPCS3 -c 1.2.3.4'
```

From OMVS shell:

```
moddvipa -p TCPCS3 -c 1.2.3.4
```

From the TSO command prompt:

```
moddvipa -p TCPCS3 -c 1.2.3.4
```

Restriction: MODDVIPA must be listed in the AUTHCMD NAMES section of the IKJTSO*xx* member of SYS1.PARMLIB.

Defining a security profile for SIOCSVIPA, SIOCSVIPA6, and MODDVIPA

You can define a System Authorization Facility (SAF) resource profile in the SERVAUTH class to control access to dynamic VIPAs (DVIPAs) in a VIPARANGE statement in the following ways:

- Control which applications can issue a SIOCSVIPA or SIOCSVIPA6 ioctl call, or call the MODDVIPA utility (which issues the SIOCSVIPA or SIOCSVIPA6 ioctl call on behalf of the application), to create a DVIPA

Guideline: Wherever SIOCSVIPA is used in this information, the information also applies to SIOCSVIPA6.

- Control whether an application can issue a SIOCSVIPA ioctl call or call the MODDVIPA utility to create a DVIPA within a specific VIPARANGE subnet

For information about controlling which applications can bind to create a DVIPA, see "Defining a security profile for binding to DVIPAs in the VIPARANGE statement" on page 401.

This information includes examples that use RACF. If you are using another security product, see the documentation for that product for the equivalent commands. For more information about RACF, see z/OS Security Server RACF Security Administrator's Guide.

Steps for controlling which applications can issue a SIOCSVIPA ioctl call or call the MODDVIPA utility to create a DVIPA

You can define a System Authorization Facility (SAF) resource profile in the SERVAUTH class to control access to dynamic VIPAs (DVIPAs) in a VIPARANGE statement.

Procedure

Perform the following steps to control whether an application can issue a SIOCSVIPA ioctl call or call the MODDVIPA utility to create a DVIPA:

1. Define the EZB.MODDVIPA.*sysname*.*tcpname* resource profile in the SERVAUTH class by issuing the following RACF command:

   ```
   RDEFINE SERVAUTH (EZB.MODDVIPA.sysname.tcpname) UACC(NONE)
   ```

2. Give the user ID that is associated with the application READ access to the resource by issuing the following RACF command:

   ```
   PERMIT EZB.MODDVIPA.sysname.tcpname  ACCESS(READ) CLASS(SERVAUTH) ID(userid)
   ```

3. Refresh the profile by issuing the following RACF command:

   ```
   SETROPTS RACLIST(SERVAUTH) REFRESH
   ```

Results

In this example, *sysname* is the name of the MVS system, *tcpname* is the job name of the TCP/IP started task, and *userid* is the user ID that is associated with the application. The job name for started tasks, such as TCP/IP, is derived depending on how it is started:

- If the START command is issued with the name of a member in a cataloged procedure library (for example, S TCPIPX), the job name will be the member name (for example, TCPIPX).
- If the member name on the START command is qualified by a started task identifier (for example, S TCPIPX.ABC), the job name will be the started task identifier (for example, ABC). The started task identifier is not visible to all MVS components, but TCP/IP uses it to build the RACF resource name.
- The JOBNAME parameter can also be used on the START command to identify the job name (for example, S TCPIPX,JOBNAME=XYZ).
- The JOBNAME can also be included on the JOB card.

Results:

- If this resource profile is defined and the user ID has READ access to the resource, then the call is processed.
- If this resource profile is not defined and the user ID is not APF authorized and is not a UNIX System Services superuser, the call fails.
- If this resource profile is defined and the user ID does not have READ access to the resource, then the call fails with a permission denied error, regardless of whether the user is a superuser.

Rules:

- If specific and wildcard resource profile names are defined, the user ID must have READ access to the resource that has the most specific match. For example, defining EZB.MODDVIPA.MVS1.TCPCS1 and EZB.MODDVIPA.MVS1.* has the following results:
  - To create a DVIPA on system MVS1, TCP/IP stack TCPCS1, the user ID must have READ access to EZB.MODDVIPA.MVS1.TCPCS1.
  - To create a DVIPA on system MVS1, TCP/IP stack TCPCS2, the user ID must have READ access to EZB.MODDVIPA.MVS1.*.
  - If a user ID tries to create a DVIPA on a TCP/IP stack on system MVS2, there is no matching resource profile.
- If this resource profile is defined, the user ID must also have READ access to the resource to delete a DVIPA that was created using the SIOCSVIPA ioctl call, the SIOCSVIPA6 ioctl call, or the MODDVIPA utility.

Steps for controlling whether an application can issue a SIOCSVIPA ioctl call or call the MODDVIPA utility to create a DVIPA within a specific VIPARANGE subnet

You can define a System Authorization Facility (SAF) resource profile in the SERVAUTH class to control access to dynamic VIPAs (DVIPAs) within a specific VIPARANGE subnet.

Procedure

Perform the following steps to protect a specific VIPARANGE subnet so that only a particular application can issue a SIOCSVIPA ioctl call or call the MODDVIPA utility to create a DVIPA in the protected range.

1. Define the EZB.MODDVIPA.*sysname*.*tcpname*.*resname* resource profile in the SERVAUTH class.

   For example, to protect a VIPARANGE subnet for application APPL1, you can define the EZB.MODDVIPA.*sysname*.*tcpname*.APPL1 resource profile using the following command:

   ```
   RDEFINE SERVAUTH (EZB.MODDVIPA.sysname.tcpname.APPL1) UACC(NONE)
   ```

2. On the VIPARANGE statement, include the SAF parameter with a *resname* value that matches the *resname* value on the EZB.MODDVIPA.*sysname*.*tcpname*.*resname* resource profile.

   For example:

   ```
   VIPARANGE DEFINE 255.255.255.255 10.10.10.1 SAF APPL1
   ```

   For more information about the SAF parameter on the VIPARANGE statement, see z/OS Communications Server: IP Configuration Reference.

3. Give the user ID that is associated with the application READ access to the EZB.MODDVIPA.*sysname*.*tcpname*.*resname* resource.

   For example:

   ```
   PERMIT EZB.MODDVIPA.sysname.tcpname.APPL1 ACCESS(READ) CLASS(SERVAUTH) ID(userid)
   ```

4. Refresh the resource profile by issuing the following RACF command:

   ```
   SETROPTS RACLIST(SERVAUTH) REFRESH
   ```

Results

In this example, *sysname* is the name of the MVS system, *tcpname* is the job name of the TCP/IP started task, and *userid* is the user ID that is associated with the application. For information about how the job name for a started task, such as TCP/IP, is determined, see “Steps for controlling which applications can issue a SIOCSVIPA ioctl call or call the MODDVIPA utility to create a DVIPA” on page 378.

Results:

- If the SAF parameter is specified and the user ID has READ access to the resource, then the call is processed.
- If the SAF parameter is specified and the user ID does not have READ access to the resource, then the call fails.
- If the SAF parameter is specified but the resource profile is not defined, then the call fails.
- All of the following profile specifications that include wildcard values match the EZB.MODDVIPA.*sysname.tcpname.resname* resource profile name:
  - EZB.MODDVIPA.*.*
  - EZB.MODDVIPA.**
  - EZB.MODDVIPA.*.*.*

  The most specific match to a resource profile name is always used. For example, consider the following VIPARANGE definitions:

```
VIPARANGE DEFINE 255.255.255.0 10.10.10.1
VIPARANGE DEFINE 255.255.255.255 10.10.10.210 SAF APPL1
VIPARANGE DEFINE 255.255.255.255 10.10.10.211 SAF APPL2
```

  If `moddvipa -p TCPCS -c 10.10.10.20` is issued, then the first VIPARANGE statement matches and EZB.MODDVIPA.*sysname.tcpname* is used.

  If `moddvipa -p TCPCS -c 10.10.10.210` is issued, although the first VIPARANGE statement matches, the most specific VIPARANGE statement match from these sample statements is the second statement, so the EZB.MODDVIPA.*sysname.tcpname*.APPL1 resource is used.

Rule: If this resource profile is defined and you want to delete a DVIPA that was created using the SIOCSVIPA ioctl call, the SIOCSVIPA6 ioctl call, or the MODDVIPA utility, the user ID must have READ access to the resource that was used to create the DVIPA.

Configuring application-instance DVIPAs for IBM z/OS Container Extensions (zCX)

IBM z/OS Container Extensions (zCX) provides an execution environment allowing z/OS to host applications based on Linux docker containers. Each instance of zCX (unique zCX job) is provisioned within a unique z/OS address space. The zCX address space represents a virtual server. The virtual server does not require operational tasks from the z/OS administrator. From an operational perspective, the virtual server is transparent to the z/OS environment. z/OS Communications Server TCP/IP provides network communications for the zCX workloads.

The zCX server (address space) is represented by a unique z/OS application instance Dynamic VIPA. Both IPv4 and IPv6 zCX DVIPAs are defined using VIPARANGE with the ZCX keyword. Each zCX address space can connect to the network using a single IPv4 DVIPA or both a single IPv4 and a single IPv6 DVIPA.

For an overview of the z/OS Communications Server networking support provided for the zCX docker environment, see “IBM z/OS Container Extensions network overview” on page 110.

Configuring the zCX server IP address

The zCX job for the docker environment must be configured using z/OSMF workflows. The zCX configuration requires defining various job related parameters, including some network related parameters which includes the virtual server's IP address. The IP address specified in the zCX configuration (workflows) must match (be coordinated with) your z/OS TCP/IP configuration for VIPARANGE ZCX. For more information about using z/OSMF workflows for zCX, see IBM z/OS Container Extensions Guide and *IBM z/OS Management Facility Online Help* for Configuration Workflow.

zCX DVIPA operational characteristics and considerations

- Dynamic VIPAs created from the VIPARANGE subnet with ZCX will not move from one stack to another stack. The MOVEable parameter is ignored.

- The value of ZCX for a VIPARANGE statement cannot be changed without first removing the existing VIPARANGE statement and then redefining it.
- The SAF parameter can be dynamically changed.
- The creation of zCX DVIPAs is protected by the same authorization functions provided for all VIPARANGE DVIPAs. The zCX instance and the associated user ID of the zCX started task must adhere the VIPARANGE authorization requirements. See "Use of the SIOCSVIPA or SIOCSVIPA6 ioctl command" on page 375 for more information. Refer to the z/OSMF workflows for configuring your zCX job.
- zCX DVIPAs are controlled by internal processing related to the zCX workload. When the zCX server is started, the DVIPA will automatically be activated and added to the home list. TCP/IP creates a static route to the zCX DVIPAs for its own internal use. This static route is local to and controlled by the stack that owns the zCX instance and it is not advertised using dynamic routing. Native z/OS applications cannot bind to a zCX DVIPA and IOCTLs cannot change the state of a zCX DVIPA.
- If IP filters are defined, updates to the IP filter rules are required for all zCX DVIPAs. Two IPSec rules are required for all of the zCX DVIPAs, one rule defining the zCX DVIPAs as the source and another rule for the zCX DVIPAs as the destination. Both rules must be defined with ROUTING EITHER, permitting both ROUTED and LOCAL traffic for all of the zCX DVIPAs. When configuring the IPSec rules through the z/OSMF Network Configuration Assistant, the topology should indicate **Filtering only**. Be sure to check both **For local traffic - Host** and **For routed traffic - Gateway** under the **Filtering only** option.

zCX DVIPA resiliency

There are usage scenarios where the zCX job (workload) will need to be relocated to another instance of z/OS. For this reason, the zCX DVIPA should be defined in all z/OS systems that could become eligible to host this zCX workload. The zCX DVIPA can only be active (used) within a single instance of z/OS at one time. However, using the z/OS Dynamic VIPA support, the zCX job could terminate (planned or unplanned) and then be restarted in another z/OS instance without reconfiguring the zCX job or z/OS TCP/IP.

When using dynamic routing, the zCX DVIPAs must be added to the OMPRoute configuration. The DVIPAs are treated by OMPROUTE as normal application-instance DVIPAs where the DVIPAs are advertised accordingly.

When using static routes you must ensure the zCX DVIPAs can be reached by other hosts by defining static routes on other hosts that need to reach the zCX instance. If a zCX DVIPA is moved to another z/OS, updates to the static route definitions will be required.

Choosing which form of dynamic VIPA support to use

The following list explains which of the features to use for the type of application that is being used.

- When to use VIPADEFINE and VIPABACKUP to define a dynamic VIPA:
  - One or more applications bind to the IPv4 INADDR_ANY address or to the IPv6 unspecified address (in6addr_any), and the applications exist on multiple TCP/IP stacks.
  - Dynamic VIPA takeover is wanted.
  - The DVIPA does not need to be deleted when the application is stopped.
- When to use VIPARANGE and bind() to define a dynamic VIPA:
  - The application cannot bind to INADDR_ANY or in6addr_any, or dynamic VIPA takeover is not wanted.
  - The IP address to which the application binds can be controlled by the user. The first explicit bind (the listening socket) of the application remains for the life of the application. Otherwise, the DVIPA is removed every time the DVIPA owning socket of the application is closed, and readded every time there is a new DVIPA owning socket (another explicit bind is done and the DVIPA does not exist).
  - Automatic deletion of the dynamic VIPA when the application is stopped is acceptable.
  - A specific dynamic VIPA address must be associated with a specific application.

  - The application is not APF-authorized, or not run under a user ID with superuser authority.
- When to use the VIPARANGE and the MODDVIPA utilities, or the SIOCSVIPA or SIOCSVIPA6 ioctl command to define a dynamic VIPA:
  - The application cannot bind to INADDR_ANY or in6addr_any, or dynamic VIPA takeover is not wanted. The IP address to which the application binds is known but cannot be controlled by the user. Automatic deletion of the dynamic VIPA when the application is stopped is not acceptable.
  - The application opens an AF_INET6 listener and binds to in6addr_any. The application needs to receive connections for multiple application-specific IPv4 and IPv6 destinations that are targeted for the port of this socket. The application can use MODDVIPA or the SIOCSVIPA and SIOCSVIPA6 ioctl command to create the IP address targets.

  The MODDVIPA utility or application that is issuing the ioctl command is to be run from an APF-authorized library and under a user ID with appropriate authority. See "Use of the SIOCSVIPA or SIOCSVIPA6 ioctl command" on page 375 or "Using the MODDVIPA utility" on page 376.

Configuring distributed DVIPAs - sysplex distributor

A distributed DVIPA exists on several stacks, but is advertised outside the sysplex by only one stack. This stack receives all incoming connection requests and routes them to all the stacks in the distribution list for processing. This provides the benefit of distributing the workload of incoming requests and providing additional fail-safe precautions in the event of a server failure.

You can distribute connections destined for a dynamic VIPA (DVIPA) by adding a VIPADISTRIBUTE configuration statement for a previously defined dynamic VIPA. The order of the statements is important. The VIPA is first defined with the VIPADEFINE statement and then included on a VIPADISTRIBUTE statement. Another TCP/IP can act as a backup for the distributed DVIPA by properly coding a VIPABACKUP statement; the backup will perform the routing function in the event of a failure. The options specified on a VIPADISTRIBUTE statement are inherited by a backup stack unless the second stack has its own VIPADISTRIBUTE statement for that DVIPA, in which case it will use its own VIPADISTRIBUTE statement for distributing. You can also code a VIPADISTRIBUTE statement with just the VIPABACKUP statement and not for the VIPADEFINE statement. This would allow workload distribution only during a primary outage.

You can change the distribution of a DVIPA after a backup stack has activated it. However, if the backup stack did not not have its own distribution defined by a VIPADISTRIBUTE statement before it activated the DVIPA, any distribution changes made while the DVIPA is active on the backup stack are temporary. Those changes will be in effect while the DVIPA remains active on the backup stack, but will not be remembered if this stack takes over the DVIPA again in the future.

The following example is a properly coded distributed DVIPA:

```
IPCONFIG SYSPLEXROUTING DYNAMICXCF 193.9.200.4 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2000::93:9:200:4
VIPADYNAMIC
  VIPADEFINE  255.255.255.192 9.67.240.02
  VIPADISTRIBUTE DEFINE  9.67.240.02 PORT 20 21 8000 9000 DESTIP
          193.9.200.2
          193.9.200.4
          193.9.200.6
  VIPADEFINE V6DVIPA1 2000::9:67:240:2
  VIPADISTRIBUTE DEFINE V6DVIPA1 PORT 20 21 8000 9000 DESTIP
          2000::93:9:200:2
          2000::93:9:200:4
          2000::93:9:200:6
ENDVIPADYNAMIC
```

Prior to z/OS V1R6 Communications Server, the TCP/IP stack that was configured as a distributor of dynamic VIPAs was required to enable IP forwarding using the IPCONFIG (or IPCONFIG6) DATAGRAMFWD TCP/IP profile statement. For installations that do not want to configure their TCP/IP stack as a forwarding node, it is no longer a requirement for distributing dynamic VIPAs. However, if your installation is configured such that target TCP/IP stacks have only XCF connectivity, datagram forwarding

still needs to be configured on the distributor, as all packets originating from the target will be forwarded by the distributor.

There are several configuration changes that can be made to affect the method the distributing stack will use to forward connections to the target stacks. In each of the following items, *all participating stacks* is used to refer to the distributing stack and all target stacks.

WLM-based forwarding based on target system workload

If the DISTMethod BASEWLM parameter is specified on the respective VIPADISTRIBUTE statement, or if the DISTMethod parameter is not specified, this distribution method is enabled. This is the default distribution method. To enable the distributing stack to forward connections based upon the workload of each of the target stacks, specify SYSPLEXROUTING on the IPCONFIG statement in all participating stacks. This registers all participating stacks with WLM and enables the distributing stack to request workload information from WLM.

The WLM workload information is based on a comparison of available general CPU capacity for each target system. If the application uses System z Application Assist Processor (zAAP) capacity or System z Integrated Information Processor (zIIP) capacity, you can configure the VIPADISTRIBUTE statement so that available zAAP CPU capacity and zIIP CPU capacity are also considered. For these additional processor types to be considered, no distributor and target systems used by this application can be earlier than z/OS V1R9 Communications Server. If you need to consider zAAP and zIIP CPU capacity, evaluate whether you can use SERVERWLM distribution as an alternative to BASEWLM distribution for this application. SERVERWLM distribution has the advantage that processor proportions are automatically determined and dynamically updated by WLM based on the actual CPU usage of the application. If you need BASEWLM distribution, to determine the processor proportions to configure, study the workload usage of assist processors by analyzing SMF records, using performance monitors reports such as RMF, and so on.

WLM-based forwarding based on server-specific workload

If the DISTMethod SERVERWLM parameter is specified on the respective VIPADISTRIBUTE statement, the distributing stack selects from the available servers for a DVIPA/port and forwards connections based on a WLM recommendation indicating how well each server is executing on its system. To enable the distributing stack to forward connections based on server-specific workload, specify SYSPLEXROUTING on the IPCONFIG statement in all participating stacks.

If the server uses System z Application Assist Processor (zAAP) capacity or System z Integrated Information Processor (zIIP) capacity, processor proportions are automatically determined and dynamically updated by WLM, based on the actual CPU usage of the application; however, you can influence the WLM server-specific recommendation with configuration options on the VIPADISTRIBUTE statement. You can use the PROCXCOST parameter on the VIPADISTRIBUTE statement so that the WLM recommendation favors servers with available zAAP or zIIP capacity over servers on which work targeted for the specialty processors might instead run on the conventional processor. You can also use the ILWEIGHTING parameter on the VIPADISTRIBUTE statement to influence how aggressively the WLM recommendation favors servers on systems with displaceable capacity at lower importance levels over servers on systems with displaceable capacity at higher importance levels. For these additional factors to be considered by WLM, no systems can run a release prior to z/OS V1R11 Communications Server.

WLM/QoS-based forwarding

Regardless of whether BASEWLM or SERVERWLM weights are being used, to enable the distributing stack to forward connections based upon a combination of workload information and network performance information (TCP retransmissions and time-outs), specify SYSPLEXROUTING on the IPCONFIG statement in all participating stacks, and also define a sysplex distributor performance policy on the target stacks. For information on configuring these policies, see "Sysplex distributor policy example" on page 840.

Round-robin forwarding

If the DISTMethod ROUNDROBIN parameter is specified on the respective VIPADISTRIBUTE statement, the distributing stack uses a round-robin mechanism to select one of the DVIPA/port targets for each connection.

Weighted active forwarding
If the DISTMethod WEIGHTEDActive parameter is specified on the respective VIPADISTRIBUTE statement, the distribution of incoming TCP connection requests is balanced across the targets, such that the number of active connections on each target is proportionally equivalent to a configured active connection weight for each target. However, server-specific abnormal completion information, server-specific health information, and the TSR value are used to reduce the active connection weight when these indicators are not optimal. To enable the distributing stack to use server-specific abnormal completion and health information to affect the active connection weight, specify SYSPLEXROUTING on the IPCONFIG statement for all participating stacks.

Hot-standby forwarding
If the DISTMethod HOTSTANDBY parameter is specified on the respective VIPADISTRIBUTE statement, one preferred target server and one or more backup (hot-standby) target servers are configured. The distributing stack does not perform load balancing of new connection requests across multiple targets; instead, the preferred target server with an active listener receives all new incoming connection requests, and the hot-standby target servers, which typically also have a ready listener application, do not receive any new connection requests. If the preferred target server becomes unavailable, then the highest ranked backup server becomes the active target server and receives all new connection requests. To enable the distributing stack to use server-specific abnormal completion and health information to affect the availability of the preferred target, specify SYSPLEXROUTING on the IPCONFIG statement for all participating stacks.

Result: If you have not made the changes needed to enable WLM-based forwarding (SYSPLEXROUTING has not been specified for all participating stacks), and BASEWLM or SERVERWLM is specified, the distributing stack will use round-robin forwarding to distribute connections.

Tip: The weights received from WLM are returned based on the first 24 characters of the HOSTNAME value. The HOSTNAME value is determined by the search path at initialization time. To ensure correct distribution, verify that the first 24 characters of each HOSTNAME value are unique for every system.

Regardless of the distribution method used, sysplex distributor routing policies can further affect the distribution of connections. Sysplex distributor routing policies, configured on the distributing stack, are used to specify a set of target stacks for a given set of traffic. For example, all traffic destined to a given port/DVIPA from a specified subnet can be assigned one group of target stacks, while traffic for the same port/DVIPA from another subnet can be assigned to a different group of target stacks. For more information on configuring these types of policies, see “Sysplex distributor policy example” on page 840.

Distribution of connections to target servers can also be affected by the responsiveness of the target stacks or target servers. The sysplex distributor stack monitors how well target servers respond to connection setup requests, calculating a target server responsiveness (TSR) value for each server.

For WLM-based forwarding based on target system workload, WLM/QoS-based forwarding, WLM-based forwarding based on server-specific workload, or weighted active forwarding, new connection setup requests are diverted away from target servers that are handling connection setup requests relatively poorly. For round-robin forwarding, if the sysplex distributor determines that a target server is not successfully accepting connection setup requests at all, that target server is bypassed. Periodically, the distributor sends a new connection request to a target with a TSR of 0, to check whether the responsiveness of that target has improved.

By default, the sysplex distributor updates the status of each target server at 1-minute intervals as follows:

- When BASEWLM distribution is being used, the distributor polls WLM for new system weights.
- When SERVERWLM distribution is being used, each target stack polls WLM for server-specific weights and sends these weights to the distributor.
- The TSR value for each server is updated.

The default 1-minute interval is sufficient for most workloads. However, in some environments, particularly when the load on each target system will be close to 100% capacity and when the workload consists of a high volume of short-lived connections, you might want to use a shorter interval so that the distributor reacts faster to changes in a target server's status. You can change the interval by using the SYSPLEXWLMPOLL parameter on the GLOBALCONFIG statement.

Each distributing stack and each target stack must have an IPv4 or IPv6 DYNAMICXCF address, or both. This address is used by other distributing stacks as a destination point. When using sysplex distributor, do not define an IUTSAMEH link. These links will be created automatically from the DYNAMICXCF statement. See z/OS Communications Server: IP Configuration Reference for directions for coding DYNAMICXCF on the IPCONFIG or IPCONFIG6 statements. For more information on additional configuration parameters required, also see the usage notes related to the DYNAMICXCF parameter under the IPCONFIG or IPCONFIG6 statements in z/OS Communications Server: IP Configuration Reference.

The VIPADISTRIBUTE statement specifies how new connection requests are routed to a set of candidate target stacks. The VIPADISTRIBUTEd DVIPA can be followed by up to 64 ports. The preceding example shows the well-known ports for FTP and the ports for a custom application.

Up to 32 target TCP/IPs follow the DESTIP keyword and are identified by their respective dynamic XCF IP addresses. Alternatively, the VIPADISTRIBUTE statement can specify DESTIP ALL, in which case all current and future stacks with activated dynamic XCF can participate in the distribution as candidate target stacks. As an application listens to one of the specified ports on each listed TCP/IP, the routing TCP/IP begins to forward connections to that stack.

Guideline: If you are using OMPROUTE for connectivity to a dynamic VIPA, the LPAR with the distributing stack that owns that dynamic VIPA is the only target stack (no remote target stacks are available), and a HiperSockets (iQDIO) interface is not configured, code a static route that represents the shared IP address in the attached network router to maintain connectivity; otherwise, because a HiperSockets interface is not configured, OMPROUTE does not advertise the routing information representing the shared IP address for the dynamic XCF interfaces in that LPAR, and the address might become unreachable because the interfaces to the remote target stacks are deleted or marked inactive. Unlike IUTSAMEH and DXCF interfaces, a HiperSockets interface that shares an IP address can remain active when there are no more remote target stacks available, and OMPROUTE advertises the routing information to the neighboring network routers (for example, Cisco) for connectivity to the shared IP address.

For more information about sysplex distributor, see *IBM z/OS V1R13 Communications Server TCP/IP Implementation, Volume 3: High Availability, Scalability, and Performance*, SG24–7998.

Manually quiescing DVIPA sysplex distributor server applications

You can stop a particular server application, or all server applications, on a target stack from receiving new DVIPA sysplex distributor workload using the VARY TCPIP,,SYSPLEX command with the QUIESCE parameter and either the PORT=*portnum*, JOBNAME=*jobname*, or TARGET parameter.

This command, entered on the target stack where the application exists, prevents the application from receiving any new DVIPA sysplex distributor connections but does not affect existing connections. This command can be issued to gracefully quiesce a target application or target system, before the application or system is brought down for planned maintenance. It can also be used to temporarily divert new workload requests from a particular application or a target stack.

The VARY TCPIP,,SYSPLEX command with the RESUME parameter and either the PORT=*portnum*, JOBNAME=*jobname*, or TARGET parameter can then be used to resume using the application as a target for new DVIPA sysplex distributor connections.

For more details on the VARY TCPIP,,SYSPLEX command and its parameters, see z/OS Communications Server: IP System Administrator's Commands.

Route selection for distributing packets

Sysplex distributor uses the dynamic XCF interfaces (DYNAMICXCF parameter on the IPCONFIG or IPCONFIG6 statements) to distribute all incoming packets to target stacks.

Using dynamic XCF interfaces has several advantages, such as simplified configuration, because you can leverage the existing XCF communication links in your sysplex environment and do not have to define separate communication paths to all target systems. This is especially useful in scenarios where target systems do not have direct network connectivity, but rather rely on the network connectivity of the routing stacks.

Sysplex distributor also includes some optimization logic that enables it to select alternative network paths for forwarding DVIPA packets. This optimization logic is automatically performed by the sysplex distributor in the following configurations:

- When the routing stack and the target stacks are in the same MVS image, the routing stack uses internal communication links (IUTSAMEH) to forward DVIPA packets to the target, avoiding an external network altogether.
- When the routing stack and a target stack are running in different LPARs but on the same System z central processor complex (CPC), and HiperSockets connectivity is available, the HiperSockets connectivity is automatically used for forwarding DVIPA packets instead of the dynamic XCF interfaces.

These optimizations provide the best performance characteristics for forwarding DVIPA packets within the same CPC.

For configurations where the routing and target stacks are in different CPCs, there are some scenarios where it might be desirable to use interfaces other than the dynamic XCF interfaces for forwarding distributed DVIPA packets:

- When sysplex XCF communication interface links, whether configured through CTCs or through the coupling facility, are constrained or are heavily used by other, non-DVIPA communications.
- When the routing and target stacks have direct network connectivity over high speed, low latency, and wide bandwidth networks, such as access to the same Gigabit Ethernet segments using the OSA-Express feature.

In these configurations, forwarding DVIPA packets over these alternative network connections can actually improve performance (that is, reduce latency and CPU cost) while reducing the usage of sysplex XCF interfaces.

You can use the VIPAROUTE statement in the VIPADYNAMIC block in the TCP/IP profile to influence the sysplex distributor's logic in selecting a route and interface to forward DVIPA packets to a target stack. Using the VIPAROUTE statement, you can indicate which IP address on the target stack is to be used as the destination or target IP address during the route lookup selection. When sysplex distributor processes the incoming packet, it determines whether a matching VIPAROUTE statement has been defined, and if it has, the TCP/IP stack returns the best available route to reach the target IP address. The incoming packet is then encapsulated using a destination of the VIPAROUTE target IP address in the outer header, and forwarded to the target stack. Generic routing encapsulation (GRE) is used to encapsulate IPv4 packets, while an outer IPv6 header is used to encapsulate IPv6 packets. This enables you to select the optimal interface for forwarding DVIPA packets across different CPCs, while retaining the optimized communication paths when the routing and target stacks are on the same MVS image, the same CPC, or both.

If you do not want to use the dynamic XCF interfaces at all, you must define the dynamic routes so that the cost to reach the IP address on the target stack using the dynamic XCF interfaces is higher than the cost to reach that IP address using other interfaces. For more information, see “Steps for configuring OSPF and RIP (IPv4 and IPv6)” on page 305. If this is not done, it is possible that dynamic XCF interfaces will be used if the normal routing tables select those interfaces. By ensuring that the route using the dynamic XCF interfaces has a higher cost, you can exploit the other interfaces for forwarding most DVIPA traffic, yet maintain the dynamic XCF interface as a backup should the preferred network interfaces fail.

In the following cases, even though VIPAROUTE was specified, the dynamic XCF interface is used for distribution:

- A target IP address is specified that is not owned by the target stack.
- The defined dynamic XCF address is for a pre-V1R7 target stack.

When these conditions are detected, messages are issued at the distributing stack, as well as when the distributing stack first attempts to route a connection request to the target stack.

The dynamic XCF address must still be configured even if VIPAROUTE definitions are used, as sysplex distributor continues to use a dynamic XCF address to identify every target TCP/IP stack in the sysplex. In addition, there are several functions that continue to depend on dynamic XCF connectivity for intra-sysplex communications:

- Sysplex-Wide Security Associations (SWSA)
- Multilevel security packets
- QoS performance data collection of Policy Agent

Generic routing encapsulation

Generic routing encapsulation (GRE) is a standard protocol described by RFC 1701, and is used for several TCP/IP functions. GRE is implemented for only IPv4 packets. GRE enables a wrapper to be placed around a packet during transmission of the data. A receiving stack that supports GRE removes the GRE wrapper, enabling the original packet to be processed by the receiving stack. GRE is often used to deliver a packet to a stack using an alternate destination IP address.

Fragmentation considerations

For IPv4, when incoming packets destined for a DVIPA address need to be forwarded to a target TCP/IP stack using a route that was determined by a VIPAROUTE statement, the packet is encapsulated using Generic routing encapsulation (GRE) prior to being forwarded. This enables the packet to be forwarded through the network to the target stack while preserving the original packet's destination IP address (that is, the DVIPA address). The GRE encapsulation process increases the size of the forwarded packet by 28 bytes. As a result, if the size of the encapsulated GRE packets are larger than the maximum transmission unit (MTU) of the network interface that will be used for forwarding the packet, the TCP/IP stack might need to perform fragmentation, creating two or more packets that are forwarded to the target stack. The target stack then reassembles the fragmented packets.

While fragmentation and reassembly processing is not unusual in an IP network, it is desirable to eliminate the need for this processing, optimizing performance. For fragmentation as a result of GRE encapsulation, the cost of the fragmentation and reassembly processing might become a concern if a large percentage of the incoming DVIPA packets to be forwarded require fragmentation. Fortunately, configuration options do exist that can help eliminate the need for this fragmentation and reassembly processing, including the following options:

- The AdjustDVIPAMSS parameter on the GlobalConfig statement defaults to **AUTO**. The value of **AUTO** indicates that TCP/IP automatically adjusts the Maximum Segment Size (MSS) on the TCP connections to compensate for the extra length of the GRE header if the connections meet the following criteria:
  - For inbound connections, if the SYN packet was forwarded by a sysplex distributor using VIPAROUTE or the local stack is also a sysplex distributor and VIPAROUTE is coded.
  - For outbound connections, if the source IP address is a distributed DVIPA.

  If the **ALL** subparameter on the AdjustDVIPAMSS is specified, TCP/IP adjusts the MSS if the source IP address is a DVIPA, whether the DVIPA is distributed or not.

  Note: Both sides of the TCP connection must support the TCP MSS option. z/OS Communication Server supports this option.
- Enable path MTU discovery on the client IP hosts (that is, client hosts sending packets to the DVIPA). This enables the client hosts to dynamically discover the smallest MTU along the path from the client to the server. In the case of forwarded DVIPA traffic, the path MTU is adjusted (reduced) by the length of the GRE header, if the addition of this header would have resulted in fragmentation being required. For IPv6, path MTU discovery is automatically enabled for all hosts, and no explicit configuration should be required.
- Ensure that the MTU size of the routes over the network interfaces that are used to forward the DVIPA packets is large enough to account for the largest client packet plus the length of the GRE header. This might be an option if the clients are connected to networks with a smaller MTU size than what is available in the network paths between the z/OS hosts (that is, the TCP/IP stack forwarding the DVIPA packets and the target TCP/IP stack). Consider the following example:
  - The majority of the client hosts are connected to the network using Fast Ethernet, and as result use an MTU of 1492.

– The route selected using the VIPAROUTE statement specifies a network path over OSA-Express Gigabit Ethernet between the two z/OS hosts. Gigabit Ethernet accommodates a much larger MTU size than Fast Ethernet (8992 versus 1492). This larger MTU size with Gigabit Ethernet is often referred to as jumbo frames.

In this configuration, the z/OS hosts can be configured to take advantage of the larger MTU size when communicating with each other over the Gigabit Ethernet network. As a result, fragmentation is avoided for these forwarded DVIPA packets, as the larger MTU easily accommodates the increased packet size resulting from the GRE encapsulation. However, it is important to ensure that this larger MTU size is used only for communications among hosts where the entire network path supports the larger MTU size. Otherwise, packets sent from the z/OS hosts using the larger MTU size might need to be fragmented as they cross network boundaries that support only lower MTU sizes. As a result, when configuring larger MTU sizes, such as jumbo frames for Gigabit Ethernet, it is also important to consider enabling path MTU discovery on the hosts using the larger MTU size. This enables these hosts (in this example, the z/OS hosts) to use the larger MTU size only where appropriate, without introducing fragmentation. For more information on specifying MTU sizes on z/OS, see "Determining the maximum transmission unit" on page 107.

Dynamic port assignment

Sysplex distributor can also react dynamically to servers binding to the distributed DVIPA and creating a listening socket, adding a port to the list of ports for which connection workload balancing will occur. If the PORT parameter is omitted from the VIPADISTRIBUTE statement for the distributed DVIPA, any server that binds to the distributed DVIPA and a nonzero port will be eligible for workload distribution. If only one server binds to the distributed DVIPA and port and establishes a listening socket, that server will get all of the work. When the second server binds to the distributed DVIPA and the same port and establishes a listening socket, it will immediately become eligible to participate in connection workload balancing. TCP/IP will not enforce a limit on the number of ports that can participate in connection workload balancing per distributed DVIPA, other than the total number of allowable ports.

Sysplex-wide source VIPA

Sysplex distributor addresses the requirement of providing to clients outside a parallel sysplex a single-IP-address appearance to application instances throughout the sysplex, and also the distribution of the incoming work among the various instances. Many applications are part of a cooperative network of applications, and the sysplex applications that serve as clients to users might also have to initiate (client-like) outbound connection requests to cooperating applications. The SOURCEVIPA feature allows applications to attain independence of any physical adapter, but SOURCEVIPA is limited to statically defined VIPAs within a stack. Different instances of the same application using sysplex distributor, and thus having a single IP address for inbound connection requests, will use different IP addresses for their outbound connection requests.

These problems are resolved by allowing a single sysplex-wide VIPA to be used as the source IP address for TCP applications and to have the sysplex stacks collaborate on assigning ephemeral ports to prevent duplicate connection 4-tuples (combination of source and destination IP addresses and ports). These solutions are provided by SYSPLEXPORTS, in conjunction with either job-specific source IP addressing or sysplex-wide source VIPAs for TCP connections.

For information on job-specific or destination-specific source IP addressing using the SRCIP statement, see z/OS Communications Server: IP Configuration Reference.

Sysplex-wide source VIPAs for TCP connections

The TCPSTACKSOURCEVIPA keyword on the IPCONFIG or IPCONFIG6 statements allows users to specify a single VIPA, static or dynamic, to be used as a source IP address for TCP applications that initiate outbound connections on that stack. TCPSTACKSOURCEVIPA is in effect only when SOURCEVIPA is enabled and an application issues a connect() without a bind(). Applications that perform an explicit bind() do not use the TCPSTACKSOURCEVIPA address, such as the following applications:

- LPQ

- LPR
- REXEC as a TSO command
- RSH
- TELNET client

In addition, applications that use the Pascal API TcpOpen() call have an explicit bind() performed for them.

Guideline: Because the SRCIP profile statement provides all of the functionality of the TCPSTACKSOURCEVIPA parameter and more granularity, consider use of the SRCIP statement instead of specifying the TCPSTACKSOURCEVIPA parameter. Specifying JOBNAME * in an SRCIP profile statement provides the same result as specifying the TCPSTACKSOURCEVIPA parameter for implicit bind scenarios, and also applies to applications that issue a bind to the IPv4 INADDR_ANY address or to the IPv6 unspecified address (in6addr_any).

Tip: The TCPSTACKSOURCEVIPA parameter overrides the VIPA IP addresses in the HOME list or the SOURCEVIPAINTERFACE specification, but the TCPSTACKSOURCEVIPA parameter can be overridden. For information about the hierarchy of various ways that the source IP address of an outbound connection is determined, see "Source IP address selection" on page 244.

If you specify TCPSTACKSOURCEVIPA and do not specify SOURCEVIPA in a profile, a warning message is issued and TCPSTACKSOURCEVIPA is not enabled. The address that is specified does not need to be active on the stack at profile processing time. For example, a valid TCPSTACKSOURCEVIPA address can be an address that is to be a target address on this stack for sysplex distribution.

If the TCPSTACKSOURCEVIPA address is not an active VIPA on the stack at the time a connect() is issued, the connect() call goes through normal source IP address selection. A warning message is issued no more than once every 5 minutes (to avoid flooding the system console), indicating an attempt to use the address that is specified in TCPSTACKSOURCEVIPA failed.

TCPSTACKSOURCEVIPA can be coded on all target stacks. The target TCPSTACKSOURCEVIPA statements can specify individual unique addresses, or can be duplicates of the addresses that are specified on the distributing stack (a target DVIPA). Specifying the same DVIPA address for TCPSTACKSOURCEVIPA on the distributor and all target stacks creates a sysplex-wide source VIPA and raises the concern for coordination of ephemeral ports across the sysplex.

For information about diagnosing sysplex-wide source VIPAs for TCP connections problems, see z/OS Communications Server: IP Diagnosis Guide.

SYSPLEXPORTS

Whenever two or more application instances use the same source IP address and initate connections to the same destination IP address and port, sysplex-wide coordination of assignment of ephemeral ports is required so that the 4-tuple for each connection remains unique. As long as the source IP address is on a single stack, this coordination is not a problem because the stack manages assignment of ephemeral ports. However, with sysplex distributor applications, multiple application instances might need to initiate connections using the same distributed DVIPA, potentially to the same destination IP address and port, so uniqueness of the connection 4-tuples cannot be guaranteed unless the stacks collaborate across the sysplex for ephemeral port assignment for distributed DVIPAs. This can be done by adding the optional SYSPLEXPORTS parameter to the VIPADISTRIBUTE statement.

You must specify the SYSPLEXPORTS parameter on the first VIPADISTRIBUTE statement processed for a particular DVIPA. SYSPLEXPORTS cannot be enabled after a DVIPA has been configured for distribution. After you have enabled SYSPLEXPORTS, you cannot disable it until you have deleted all distribution for the DVIPA. If you want to specify the SYSPLEXPORTS parameter on a VIPADISTRIBUTE statement in the data set referenced by a VARY TCPIP,,OBEYFILE command, you must delete the existing VIPADISTRIBUTE statement in one data set, and then add the statement back in a second data set. An example of the contents of a data set that deletes an existing distributed DVIPA is as follows:

```
VIPADYNAMIC
VIPADISTRIBUTE DELETE
10.1.1.1 PORT 20 21
```

```
DESTIP ALL
ENDVIPADYNAMIC
;
```

An example of the contents of a data set that adds SYSPLEXPORTS to this DVIPA is as follows:

```
VIPADYNAMIC
VIPADISTRIBUTE DEFINE SYSPLEXPORTS
10.1.1.1 PORT 20 21
DESTIP ALL
;
ENDVIPADYNAMIC
```

When a distributed DVIPA can be active on more than one target stack, SYSPLEXPORTS can be specified to cause the stacks to collaborate in the assignment of ephemeral ports for outbound initiated TCP connections. This ensures that two different connections do not end up with the same connection 4-tuple.

At profile processing time, a stack with a profile that contains a SYSPLEXPORTS parameter on a VIPADISTRIBUTE statement connects to the coupling facility SYSPLEXPORTS structure containing sysplex port assignment information. The name of this structure is in the format EZBEPORT*vvtt*, where *vv* is the 2-character VTAM group ID suffix specified on the XCFGRPID start option, and *tt* is the TCP group ID suffix specified on the GLOBALCONFIG statement in the TCP/IP profile. If no VTAM group ID suffix is specified, but a TCP/IP group ID suffix is specified, *vv* is 01. If no TCP/IP group ID suffix is specified, but a VTAM group ID suffix is specified, *tt* is not present. If neither group ID suffix is specified, both *vv* and *tt* are not present. The structure will be a list structure with an entry for each DVIPA address that has a VIPADISTRIBUTE statement with SYSPLEXPORTS specified anywhere in the sysplex (or within the subplex, if subplexing is being used). The first stack to connect to the EZBEPORT*vvtt* structure for a particular DVIPA creates an entry for that DVIPA in the coupling facility. The stack creates and initializes, in the EZBEPORT*vvtt* structure, a sublist for this DVIPA of assigned ports for this stack. For more information about setting up EZBEPORT*vvtt*, see Setting up the sysplex environment for VTAM and TCP/IP functions in z/OS Communications Server: SNA Network Implementation Guide.

The stack also maintains a list of allowable ephemeral ports on this stack. These ports are not reserved for TCP by a PORT or PORTRANGE statement, are not reserved in a GLOBALCONFIG EXPLICITBINDPORTRANGE statement, and are in the range provided on the EPHEMERALPORTS parameter of the TCPCONFIG statement. The default ephemeral port range is 1024 – 65535. Only port number values in the list maintained by the stack are allocated for use by the SYSPLEXPORTS DVIPAs of the stack. Because this list is unique to a particular stack and determined by stack configuration, a port number that is not permissible for one stack because it is reserved might be allowable for another stack, and could actually be allocated for use by that stack for a SYSPLEXPORTS DVIPA.

Under the following conditions, the first time that an application issues a TCP bind() with port 0 or a connect() request, TCP/IP receives an unassigned group of ports from the coupling facility structure that are allowable as ephemeral ports on the stack (not otherwise reserved by PORT or PORTRANGE or limited by the EPHEMERALPORTS parameter):

- The bind() or connect() request uses a distributed DVIPA as the source address (whether by the application explicitly binding the socket to the designated DVIPA or by the stack assigning the TCPSTACKSOURCEVIPA address).
- The distributed DVIPA is designated as SYSPLEXPORTS.

The stack will assign an ephemeral port from the obtained group as the source port for the TCP connection request, and the coupling facility structure will be updated to show the group of ports as assigned. The stack will assign ports from within this group for each subsequent equivalent bind() request. If all the ports within the group are used, the stack will obtain another group of ports from the coupling facility.

Guideline: The maximum number of simultaneously active outbound connections that use sysplex-wide ephemeral port assignment is affected by the following parameters just like ephemeral port assignment within a single stack:

- The EPHEMERALPORTS parameter on the TCPCONFIG statement
- The EXPLICITBINDPORTRANGE parameter on the GLOBALCONFIG statement

- The PORT and PORTRANGE statements

That is, a single z/OS TCP stack supports no more than about 63000 simultaneously active, locally initiated TCP connections whose source ports are ephemeral ports assigned by the stack. If a stack is unable to successfully obtain an ephemeral port from the coupling facility for a SYSPLEXPORTS DVIPA, the connection request is terminated with an error indication.

If you send a connection request to a distributed DVIPA that is enabled for SYSPLEXPORTS and a random ephemeral port with no associated listener, then this connection will time out.

For information on diagnosing SYSPLEXPORTS problems, see z/OS Communications Server: IP Diagnosis Guide.

GLOBALCONFIG EXPLICITBINDPORTRANGE

You must specify the EXPLICITBINDPORTRANGE parameter on the GLOBALCONFIG profile statement if you are using or intend to use either of the following configurations:

- A distributed DVIPA as a source IP address in a DESTINATION rule in a SRCIP block
- A distributed DVIPA as a source IP address in a JOBNAME rule in a SRCIP block for an application that might use an IPv6 socket to connect to a mapped IPv4 destination address (this applies only if the stack is IPv6 enabled)

The EXPLICITBINDPORTRANGE parameter on the GLOBALCONFIG statement establishes a pool of ephemeral ports that is managed to guarantee the uniqueness of an assigned port across the sysplex (or subplex). This pool of ports is used to provide a sysplex-wide unique ephemeral port for any application that issues an explicit bind() to INADDR_ANY or the IPv6 unspecified address (in6addr_any) and port 0 before issuing a connect() request.

If an application uses an explicit bind() to INADDR_ANY or in6addr_any and port 0 before connecting, TCP assigns an ephemeral port from this pool to assure uniqueness across the sysplex, because it cannot be determined at the time of the explicit bind() whether the source IP address determined at connect() time might be a distributed DVIPA derived from a match on a JOBNAME or DESTINATION rule in the SRCIP block. Sysplex-wide coordination of the assigned ephemeral port value is required when the source IP address is a distributed DVIPA, so that each connection 4-tuple remains unique across the sysplex.

Restriction: The use of the EXPLICITBINDPORTRANGE pool of sysplex-wide unique ports is not supported in some common INET (CINET) configurations. It is supported if stack affinity is established, or if there is only one active TCP/IP stack for CINET to manage. In other cases, the parameter is accepted, but the results are unpredictable.

When EXPLICITBINDPORTRANGE is configured, selecting a distributed DVIPA during connect processing from a matching SRCIP JOBNAME or DESTINATION rule is also permitted when the application has explicitly bound the source port to either a port number less than 1024 or to an ephemeral port that is reserved for the job by a PORT or PORTRANGE profile statement. However, if the source port is bound to an ephemeral port that is not reserved for the job and the source address selected from a SRCIP rule during connect processing is a distributed DVIPA, the connect request will fail.

Rule: If the source port is either less than 1024 or a port that is reserved for this job and the specified source is a distributed DVIPA, ensure that multiple outbound connections to the same destination IP address and port cannot occur concurrently with the same source IP address and port.

When a profile is processed, a stack with a profile that contains a GLOBALCONFIG statement with an EXPLICITBINDPORTRANGE parameter takes the following actions:

- Connects to the coupling facility SYSPLEXPORTS structure that contains sysplex port assignment information
- Associates the stack with the pool established by the EXPLICITBINDPORTRANGE parameter

The structure is a list structure with an entry for each DVIPA address that has a VIPADISTRIBUTE statement with the SYSPLEXPORTS parameter anywhere in the sysplex (or within the subplex, if subplexing is being used). List 0 contains entries to maintain the explicit bind port range pool. Each stack

that associates itself with the pool sets the port range for that pool. The stack creates and initializes, in the structure, a sublist in list 0 of assigned ports for that stack.

Note: The name of this structure is in the format EZBEPORT*vvtt*, where the *vv* value is the 2-character VTAM group ID suffix specified on the XCFGRPID start option, and the *tt* value is the TCP group ID suffix specified on the GLOBALCONFIG statement in the TCP/IP profile. If no VTAM group ID suffix is specified, but a TCP/IP group ID suffix is specified, *vv* is 01. If no TCP/IP group ID suffix is specified, but a VTAM group ID suffix is specified, *tt* is not present. If neither group ID suffix is specified, both *vv* and *tt* are not present. For more information about setting up EZBEPORT*vvtt*, see setting up the sysplex environment for VTAM and TCP/IP functions in z/OS Communications Server: SNA Network Implementation Guide.

Guidelines:

- When the GLOBALCONFIG EXPLICITBINDPORTRANGE statement in the TCP/IP profile is processed, each stack sets the range for the pool established by the EXPLICITBINDPORTRANGE parameter for all the participating stacks in the sysplex. If each stack defines a different range, the last EXPLICITBINDPORTRANGE parameter processed supercedes any previously specified port ranges. To prevent the range from varying based on the order in which stacks complete their profile processing, specify the same port range for each stack that configures GLOBALCONFIG EXPLICITBINDPORTRANGE. You can do this by specifying the GLOBALCONFIG EXPLICITBINDPORTRANGE statement in a file that is included in each stack's TCP/IP profile using an INCLUDE statement.
- Avoid overlap of the explicit bind port range with other ports or port ranges reserved through either the PORT or PORTRANGE profile statements. If you are using the EPHEMERALPORTS parameter of the TCPCONFIG profile statement, ensure that the explicit bind port range does not consume too many of the ephemeral ports that the stack might need for general use. If a port in the pool established by the EXPLICITBINDPORTRANGE parameter is reserved on a TCP/IP stack through the PORT or PORTRANGE statements, the port cannot be allocated to that stack for EXPLICITBINDPORTRANGE processing, which might restrict the number of ports in the explicit bind port range that are available for EXPLICITBINDPORTRANGE processing. For example, if you defined an explicit bind port range 40000 – 41000, and you have a PORTRANGE statement on that stack that reserves ports 40500 – 41000, only ports 40000 – 40499 are available to that stack for EXPLICITBINDPORTRANGE processing.

  If an application binds to a port in the explicit bind port range that is reserved using a PORT or PORTRANGE statement, the bind is allowed, but the port is not treated as an EXPLICITBINDPORTRANGE port. It is not marked in the SYSPLEXPORTS coupling facility structure for coordination across the sysplex. This situation might cause connections to fail as a result of duplicate connection 4-tuples.

  If an application binds to a port in the explicit bind port range that is not reserved using a PORT or PORTRANGE statement, the bind fails with errnojr JrExpBndPortRangeConflict (x734C).

  Ports that are specified by the EXPLICITBINDPORTRANGE parameter cannot be used for normal ephemeral port allocation. Therefore, if the range of ports specified by the EXPLICITBINDPORTRANGE parameter overlaps the range of ports specified in the EPHEMERALPORTS parameter, this overlap might restrict the number of ports in the ephemeral port range. For example, if you defined an explicit bind port range 40000 – 41000, and you have an EPHEMERALPORTS parameter on the TCPCONFIG statement on that stack that limits ephemeral ports to 40000 – 41500, only ports 41001 – 41500 are available to that stack for general ephemeral port processing.

  Result: EXPLICITBINDPORTRANGE takes precedence over TCPCONFIG EPHEMERALPORTS, which means that if TCPCONFIG EPHEMERALPORTS is in use, EXPLICITBINDPORTANGE might specify ports outside the EPHEMERALPORTS range.

To remove a stack's participation in the EXPLICITBINDPORTRANGE pool, specify the GLOBALCONFIG statement with the NOEXPLICITBINDPORTRANGE parameter using a VARY TCPIP,,OBEYFILE command. This causes that stack to disassociate itself from the EXPLICITBINDPORTRANGE pool. If you do this and you have applications that issue an explicit bind() to INADDR_ANY or in6addr_any and port 0 on this stack that might match the following rules, you must ensure that no distributed DVIPAs are used as the source IP address on the matching rule.

- The DESTINATION rule of the SRCIP block on this stack

- A JOBNAME rule of the SRCIP block on this stack, where the application associated with the job name might use an IPv6 socket to connect to an IPv4 mapped destination address (this applies only to stacks that are IPv6 enabled)

Timed affinities

Sysplex distributor normally distributes each connection request, as it arrives to one of the candidate server instances, based on available capacity and policies in effect when the connection request arrives. In particular, each connection is assumed to be independent of all other existing and future connections, with respect to the server instance that will receive the incoming connection request.

Some applications, however, establish an affinity between a client and a particular server instance that needs to span multiple connections. TN3270E Telnet server printer sessions, for example, are based on a connection request from a Telnet client, and that printer session connection request needs to be routed to the same TN3270E Telnet server that is serving the LU2 session. Similarly, web-based applications, such as shopping carts, might need to have all connections from a particular client come to the server instance that has the contents of the shopping cart stored as session state.

The TIMEDAFFINITY parameter on the VIPADISTRIBUTE statement indicates to sysplex distributor that connections to a particular distributed DVIPA need to take into account the client origin. Connections from the same client, as identified by IP address, need to be routed to the same server instance, even when multiple server instances are hosted by a single target stack.

If a nonzero value for the TIMEDAFFINITY parameter is specified on a VIPADISTRIBUTE statement, the first connection from a particular client is routed as normal to a target stack and listening application. At that time, both the sysplex distributor routing stack and the target stack establish an affinity to govern subsequent connection requests from the same client. This affinity maintains a connection count, initially one. As subsequent connection requests for the same distributed DVIPA and port come in from the same client IP address, they are routed to the same server instance and the affinity connection count is incremented. As affinity-based connections, including the first one, are closed, the connection count is decremented.

When the last existing connection is closed and the count gets to zero, a timer of the duration (in seconds) specified by the TIMEDAFFINITY parameter is started. If another connection request is received from the same client to the same distributed DVIPA and port, the timer is stopped and the connection request is routed to the designated server instance. If no connection request is received from that client for the designated distributed DVIPA and port before the timer expires, the affinity is removed from the sysplex distributor routing and target stacks. The next connection request from that client for the distributed DVIPA and port will be routed according to normal sysplex distributor considerations of relative capacity and policies.

Connection requests that map to an existing affinity are always routed to that target regardless of the distribution method being used. This also applies to weighted active connection distribution. Although the established affinities are honored, distribution decisions for connection requests that have no affinity will include the active connections that were established as a result of an affinity relationship.

Using the VARY TCPIP,,OBEYFILE command with a profile containing a VIPADISTRIBUTE statement that specifies the TIMEDAFFINITY value as 0 prevents new affinities from being established. However, an existing affinity remains until all connections using the affinity end and the timer for the affinity expires. The timer value is not changed for an existing affinity by the new statement; instead, it continues to use the configured TIMEDAFFINITY value that was active when the affinity was established. To remove existing affinities, issue the VARY TCPIP,,OBEYFILE command with a profile containing a VIPADISTRIBUTE DELETE statement, followed by a VIPADISTRIBUTE DEFINE statement with a TIMEDAFFINITY value of 0.

Note that under some circumstances, a client's affinity with a specific target application server instance might be terminated prior to the specified time interval if key resources needed to satisfy new client TCP connection requests are not available. This includes the following scenarios:

- A target application server instance terminates, is quiesced for DVIPA sysplex distributor workload balancing, or is no longer listening on its specified port. Any affinities to the target application instance are terminated and new connections for that port are no longer routed to this server.

One exception to this scenario occurs for configurations where multiple applications are on the same system and TCP/IP stack, and also listen on the same port (that is, the set of applications comprises a shareport group). In this case, when one of the application instances in the shareport group terminates, is quiesced for DVIPA sysplex distributor workload balancing, or is no longer listening on the specified port, the routing stack continues to maintain any existing affinities, but only to this target system and TCP/IP stack, as long as at least one application instance in the shareport group is active, is not quiesced, and is listening to the specified port. When a new TCP connection is received from a client that previously had an affinity to the server that is no longer active, the request is routed to the same target system and TCP/IP stack. One of the other available servers in the shareport group is selected to process the request and a new affinity is established.

- A target system or TCP/IP stack is no longer available. All affinities associated with applications running on that system and TCP/IP stack are terminated.
- A target TCP/IP stack is no longer reachable across dynamic XCF links (that is, the dynamic XCF link to the target stack is no longer active). Any existing client affinities to applications on that target stack are terminated if new connection requests from these clients are received while the dynamic XCF link is not active. Note that this scenario occurs only if a dynamic XCF link becomes inactive and all attempts to automatically restart the link are unsuccessful.
- A target server is found to be unresponsive to accepting new connection requests. All affinities to the target server are terminated and new connections for that port are no longer routed to that server.
- A target server is unable to support new connections requests. Connections that are protected by IPSec UDP-encapsulated security associations that are negotiated with a peer behind a NAPT must be distributed to a V1R8 or later target. It is not always possible for the distributing stack to detect that a security association is being negotiated with a peer behind a NAPT. If an initial connection is distributed to a V1R7 target and an affinity to that target is established, and during negotiation of a subsequent security association it is determined that the peer is behind a NAPT, the affinity to the target is terminated.

The sysplex distributors notion of *client* is tied exclusively to source IP address on the connection request. This notion is not new, and is also true of other functions such as Policy Agent and IPSec. However, it can present problems in situations where many different connections from truly different client instances all appear with the same source IP address. Examples include the following situations:

- Proxy applications, such as a web HTTP server proxy, that initiate secondary connections on behalf of a large number of different clients. The secondary connections into sysplex distributor all look as though they come from the same client, and any affinity for those connections directs all connection requests to the same server.
- Network address translation (NAT), which keeps connection tables so that it can map multiple client-side addresses into a single server-side source IP address.
- Clients on z/OS configured for SOURCEVIPA, or in general, many instances of a client on the same network node with one or very few IP addresses (as far as outbound connection requests are concerned).
- Clients running on the sysplex distributor routing stack, which is a special instance of the previous scenario. The source IP address for a client running on the sysplex distributor routing node is the distributed DVIPA, the same as the destination IP address. This is not new, and is true for any connection request issued for a server on the same stack, if the connection request socket was not bound to a particular IP address before issuing the connect() call.

In cases like this, distribution might be significantly less than optimal, because sysplex distributor has no way to distinguish among connection requests from different real clients all using the same source IP address. In the worst case, where all source IP addresses are the same (for example, where there is a single proxy instance in a firewall), there will be no load balancing at all as long as an affinity exists.

Essentially, timed affinity was created to allow connection workload distribution where it was not possible before, due to the client/server application model requiring a particular client to go to a particular server instance for some period of time. Such applications are not, at present, workload balanced. If a network configuration is such that a reasonable number of source IP addresses do not allow workload balancing, techniques to partition the work at the source must be implemented (configuring the clients to go to

unique server addresses, or configuring the proxy to distribute the workload among multiple servers, as the WebSphere HTTP server plug-in does). Where the anticipated number of different source IP addresses, and the connection request arrival rate, is large enough to provide reasonable balancing and reaction to changing sysplex workloads, this provides a reasonable solution while respecting affinities between clients and servers as required for the application. The main reason for configuring the affinity on the basis of each distributed DVIPA and port pair is that affinity requirements differ from application (port number) to application, and some applications do not need affinity at all. You must take into account your specific network configuration, and specifically the arrival rate of connections from different IP addresses, in determining whether timed affinity with sysplex distributor is appropriate for a particular application in your network.

Affinity information needs to be handled on the sysplex distributor routing stack, backup stacks, and target stacks. If a target stack is not z/OS V1R5 or later, server application instances that are part of a shareport group on that stack will not work properly, and affinities for that target stack will not be communicated to the backup routing stack on failure of a primary routing stack. If a backup routing stack is not z/OS V1R5 or later, affinity information will not be sent to it from surviving target stacks if it takes over from a failed routing stack. Therefore, it is strongly recommended that all TCP/IP stacks participating in distribution for a distributed DVIPA with affinities be at least z/OS V1R5.

You can use the Netstat VCRT/-V report option with the DETAIL modifier to display the affinity related information for each connection, as follows:

```
MVS TCP/IP NETSTAT CS V1R8       TCPIP Name: TCPCS           15:10:28
Dynamic VIPA Connection Routing Table:
Dest:      203.1.10.18..21                (1)
  Source:  193.10.1.118..0
  DestXCF: 193.1.1.108
    CfgTimAff: 0200  TimAffCnt: 0000000003  TimAffLft: 0000
Dest:      203.1.10.18..21                (2)
  Source:  193.10.1.118..1026
  DestXCF: 193.1.1.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21                (2)
  Source:  193.10.1.118..1027
  DestXCF: 193.1.1.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21                (2)
  Source:  193.10.1.118..1028
  DestXCF: 193.1.1.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21                (3)
  Source:  193.10.1.119..0
  DestXCF: 193.1.2.108
    CfgTimAff: 0200  TimAffCnt: 0000000001  TimAffLft: 0000
Dest:      203.1.10.18..21                (4)
  Source:  193.10.1.119..1030
  DestXCF: 193.1.2.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      203.1.10.18..21                (5)
  Source:  193.10.1.120..0
  DestXCF: 193.1.1.108
    CfgTimAff: 0200  TimAffCnt: 0000000000  TimAffLft: 0099
Dest:      204.2.10.11..21                (6)
  Source:  193.10.1.199..1031
  DestXCF: 193.1.6.108
    PolicyRule:    FTPD1
    PolicyAction:  paPRD-SD-7-INTR-SPECIAL
Dest:      205.2.10.11..21                (6)
  Source:  193.10.1.199..1032
  DestXCF: 193.1.6.108
    PolicyRule:    *NONE*
    PolicyAction:  *NONE*
```

The following notes apply to the preceding example.

Notes:

1. Affinity CRT entry for the three regular CRT entries that follow. If there is an affinity entry, it is shown before the regular CRT entries.
2. Three regular CRT entries, associated with the single preceding affinity entry.
3. Affinity CRT entry for the regular CRT entry that follows.
4. Regular CRT entry associated with the previous affinity entry.
5. An affinity CRT entry that has no connection associated with it. The use count is zero. There are 99 seconds affinity time left before this affinity entry is removed.
6. Regular CRT entry. There is no affinity associated with it.

Sysplex-Wide Security Associations

To enable Sysplex-Wide Security Associations (SWSA) on a stack that has IP security enabled, add the DVIPSEC parameter in the IPSEC statement block of the TCP/IP profile.

To take advantage of the functions described here, you must add the DVIPSEC parameter to the primary stack that owns a DVIPA and to all backup TCP/IP stacks. It is not necessary to add the DVIPSEC parameter to hosts that serve only as targets for sysplex distributor.

You should add DVLOCALFLTR, a DVIPSEC subparameter, to the IPSEC statement block of the TCP/IP profile when intra-sysplex traffic should be protected by security associations. The DVLOCALFLTR parameter enables IP filtering and IPSec protection of TCP traffic between a client and an IPv4 dynamic VIPA that are defined on the same TCP/IP stack, when the traffic is forwarded to another TCP/IP stack.

Restriction: A security association cannot be negotiated if the source and destination IP addresses for client connections are the same dynamic VIPA. To help avoid this scenario, do not code a dynamic VIPA that can be used as a destination IP address on the TCPCONFIG TCPSTACKSOURCEVIPA or SRCIP statement in the TCP/IP profile.

For more information about configuring SWSA, see "Sysplex-Wide Security Associations and IP security" on page 1079 in this document, and IPSEC statement in z/OS Communications Server: IP Configuration Reference.

SWSA also requires the use of a coupling facility structure with a name in the form EZBDVIPA*vvtt*, where *vv* is the 2-digit VTAM group ID suffix specified on the XCFGRPID start option, and *tt* is the TCP group ID suffix specified on the GLOBALCONFIG statement in the TCP/IP profile. If no VTAM group ID suffix is specified, but a TCP/IP group ID suffix is specified, *vv* is 01. If no TCP/IP group ID suffix is specified, but a VTAM group ID suffix is specified, *tt* is not present. If neither group ID suffix is specified, both *vv* and *tt* are not present. For information about setting up the sysplex environment and the use of the EZBDVIPA*vvtt* coupling facility structure, see z/OS Communications Server: SNA Network Implementation Guide.

Dynamic IPSec security associations (SA), negotiated by IKE, can use a DVIPA address as the SA endpoint. Manually configured SAs are not supported by SWSA. For more information on IPSec, see Chapter 17, "IP security," on page 883.

When using SWSA, there are two possible configurations to consider:

- DVIPA takeover
- Sysplex distributor

To support IPSec in conjunction with DVIPA takeover and sysplex distributor, some IKE and IPSec configuration is required. Loss of access to the coupling facility is also discussed in the following subtopics.

For information on diagnosing SWSA problems, see z/OS Communications Server: IP Diagnosis Guide.

DVIPA takeover

When a DVIPA is moved during a DVIPA takeover (planned or unplanned), new Sysplex-Wide Security Associations (SWSAs) are automatically reestablished with the same security service characteristics as the Security Associations (SAs) that existed on the host that previously owned the DVIPA. When you are using IKEv1 and the SA is reestablished, the process is transparent to the client that owns the other

end of the SA. The process appears to be a normal SA refresh when IKEv1 is used to negotiate the SA. When IKEv2 is used, the SA reestablishment appears as a new SA. Figure 53 on page 397 shows DVIPA 192.168.253.4 being taken over by the backup host; SAs are transparently reestablished between the client and the backup host.

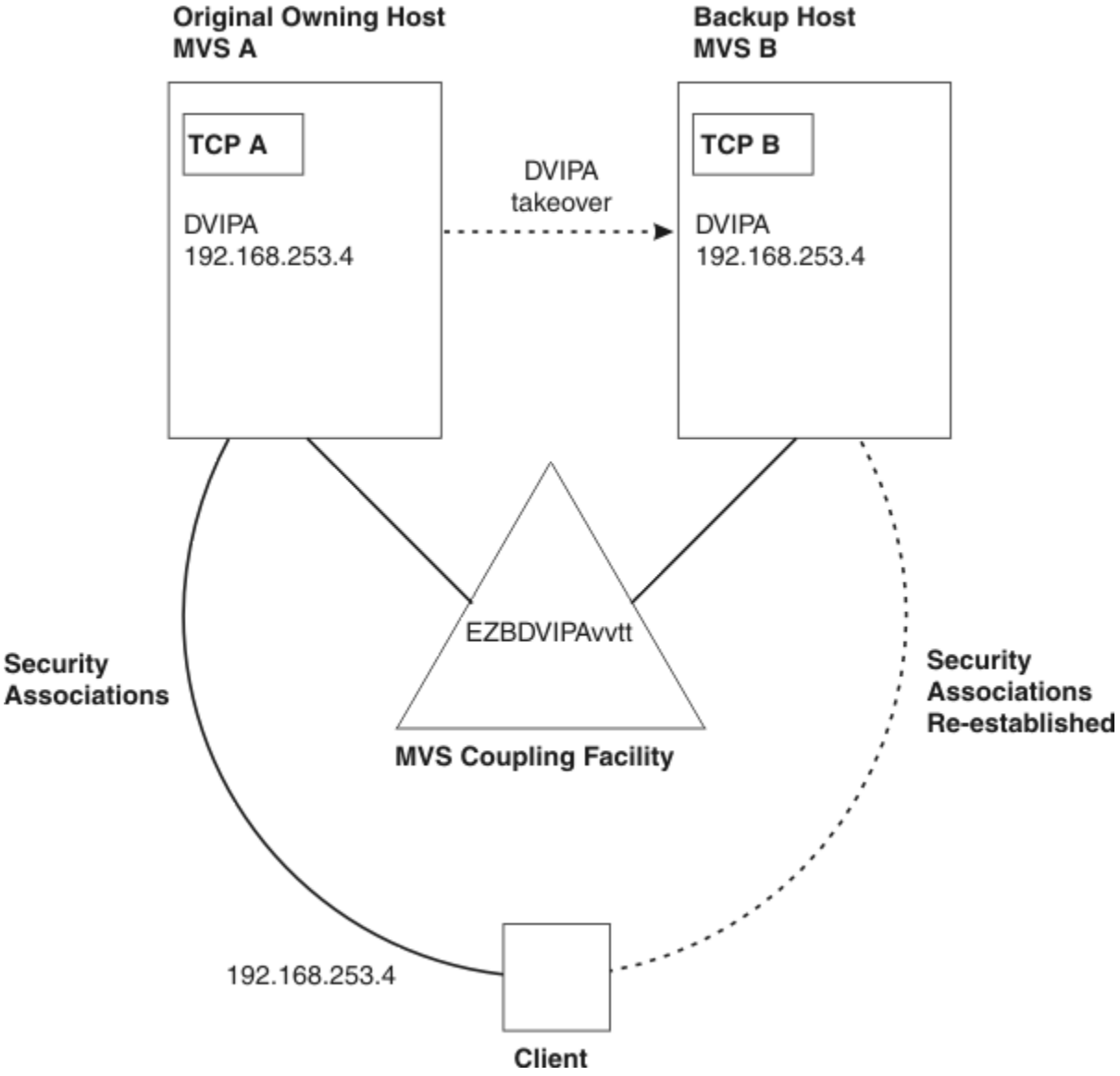


Figure 53. DVIPA takeover with SWSA

The IKE running on behalf of the TCP stack of the DVIPA owner is responsible for all IKE SA negotiations. The TCP stack owning the DVIPA is responsible for keeping the coupling facility updated with information needed to reestablish the SAs in the event of a DVIPA takeover. When a takeover occurs, the IKE on the backup host assumes responsibility for renegotiating new SAs based on information read from the coupling facility by the TCP stack of the new DVIPA owner.

When the SA is reestablished using the IKEv2 protocol, it appears as a new SA. The old SA on the client system might not be deleted immediately. This typically happens when a phase 1 SA exists on the backup system that has the same local and remote identities as the phase 1 SA protecting the SA that is being reestablished. When this scenario occurs, the old SA on the client remains active until it is deleted by liveness checking, which is described in RFC 5996. For more information about liveness checking as implemented on the z/OS platform, see the LivenessInterval parameter on the KeyExchangePolicy statement in z/OS Communications Server: IP Configuration Reference. For more information about the IKEv1 and IKEv2 protocols, see the IKE protocol details appendix in z/OS Communications Server: IP Diagnosis Guide.

Sysplex distributor

TCP traffic protected by an IPSec SA with a sysplex-distributed DVIPA endpoint can be distributed to target hosts. IPSec cryptography for inbound traffic is performed on the target host whenever possible. If not possible, the distributor performs the cryptography before forwarding the packet to the target stack. IPSec cryptography for outbound traffic is performed on the target host, and then sent directly into the network without being routed through the distributor. Figure 54 on page 398 shows the target stack performing the cryptography for the inbound and outbound traffic.

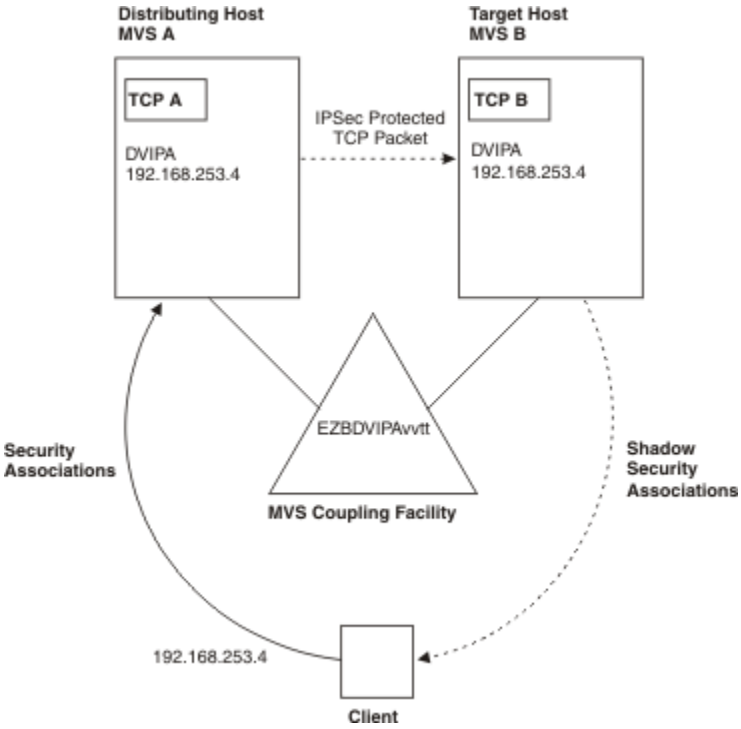


Figure 54. Sysplex distributor with SWSA

The IKE running on behalf of the distributor TCP stack (the DVIPA owner) is responsible for all IKE SA negotiations. The distributor stack keeps the master copy of the SA associated with the DVIPA. Whenever a new SA is negotiated or refreshed and the SA is installed in the distributor stack, a copy (shadow) of the SA, which contains information necessary to perform IPSec cryptography, is sent within the sysplex to the target hosts. The shadow SAs enable the distribution of cryptography to the target stacks. The coupling facility is used as a central repository for SA replay protection sequence numbers used for outbound operations. The SA lifesizes (bytes sent and received over an SA) are maintained in the master SA.

Using IPSec with DVIPAs and sysplex distributor

To support IPSec with DVIPA takeover and sysplex distributor, some IKE and IPSec configuration on the primary or distributing host must be replicated onto all systems that can either serve as a backup host for a VIPA takeover or a target host for sysplex distributor. This configuration includes IP Security policy that affects traffic using distributed DVIPA (from an IKE definition perspective).

- From a stack perspective, all anchor rules that are applicable to distributed DVIPA traffic must be identical on all systems. In addition, the ordering of the rules must allow for consistent application of security policy on all systems.
- To be considered a sysplex-wide SA, the SA negotiated that applies to DVIPAs must be at a granularity no coarser than host for the local address. That is, a dynamic SA cannot use a subnet or range that encompasses a DVIPA address. This rule ensures that on a DVIPA Giveback the SA can be moved from host to host without concerns about an SA being applicable to both the backup and primary host simultaneously. If such a dynamic SA is negotiated, the IPSec traffic using it cannot be distributed or recovered through the DVIPA takeover support.

- The configured IKE identity for a sysplex-wide SA must be identical on the primary and backup hosts. This configuration allows the backup host for a VIPA takeover to retrieve data from the coupling facility and renegotiate the SA.

Loss of access to coupling facility

If access is lost to the coupling facility that contains the DVIPA structure EZBDVIPA*vvtt*, it is possible the TCP connections that are using this DVIPA might stop and new connections that need IPsec fail to establish. Loss of access can be caused by any of the following events:

- A disconnect from the coupling facility structure.
- The structure is rebuilt.
- The structure encounters a critical storage shortage.

Loss of coupling facility access should affect only sysplex distributor connections that are being encrypted or authenticated. When access to EZBDVIPA*vvtt* is restored, the sessions can be reestablished.

If a list is not available in the EZBDVIPA structure for a DVIPA or a security association sequence number, it is possible new connections that need IPSec fail to establish. See Sysplex-wide security associations in z/OS Communications Server: SNA Network Implementation Guide for more information about increasing the number of lists for the EZBDVIPA structure.

Resolution of dynamic VIPA conflicts

The same dynamic VIPA can exist on more than one stack in the sysplex, playing different roles on the different stacks. The TCP/IP stacks collaborate to prevent conflicting definitions. For example, at any given time only one stack will advertise a given dynamic VIPA to the routers.

Potentially conflicting dynamic VIPA definitions can arise during profile processing or as the result of changes within the sysplex due to a stack or application failure or as the result of movement of workload to a different stack. The following scenarios are examples of dynamic VIPA conflict resolution handled automatically by the TCP/IP stacks. For a summary of dynamic VIPA conflict identification and resolution, see “Dynamic VIPA creation results” on page 405.

Restart of the original VIPADEFINE TCP/IP after an outage

When a dynamic VIPA is defined using VIPADEFINE on one TCP/IP, and other stacks are designated as backup using VIPABACKUP statements for the same dynamic VIPA, the stack with the highest backup rank for that DVIPA will activate it if or when the VIPADEFINE stack fails.

If the failed stack is later restarted with the same VIPADEFINE profile statement, it is likely that connections to that DVIPA will exist on the backup stack that now has the DVIPA activated and advertised to the routers. How and when ownership of the DVIPA is returned to the restarted stack is determined by how the DVIPA was originally configured.

VIPADEFINE MOVEABLE IMMEDIATE

If the DVIPA is an IPv6 DVIPA, or is an IPv4 DVIPA that was originally configured with MOVEABLE IMMEDIATE, the following events occur:

- The DVIPA ownership is immediately transferred to the restarting stack which adds the DVIPA to its HOME list and the routers are dynamically notified. The restarted stack receives all new connections for that DVIPA. The stack also can receive packets for existing connections, and it routes these to the backup stack to preserve those connections.
- At the same time, the backup stack notifies the routers that it no longer is the owner of the DVIPA.
  - If there are no current connections to the DVIPA, it is removed from the HOME list on the backup stack and it reverts to backup status.

– If there are any existing connections, the DVIPA remains in the HOME list of the backup stack and the DVIPA is put into `Moving` status until the last existing connection is terminated. At that time, the DVIPA is removed from the HOME list and reverts to backup status.

IBM recommends this form of a planned DVIPA takeback occur only during low periods of connection activity. This gives the attached routers time to update their routing tables and avoid connections being reset due to receiving an ICMP_HOST_UNREACH or ICMP6_DST_UNREACH from the router.

Tip: If OMPROUTE is used, it is recommended that GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN be configured. This causes DVIPA takeback to be delayed until OMPROUTE is active and able to advertise DVIPAs on the takeback stack. For more information on using DELAYJOIN, see "Sysplex problem detection and recovery" on page 452.

Notes:

1. To ensure preservation of existing connections on the prior owning stack, you must define DYNAMICXCF on both stacks, on the IPCONFIG statement for IPv4 dynamic VIPAs or on the IPCONFIG6 statement for IPv6 dynamic VIPAs.
2. MOVEABLE IMMEDIATE is the default for IPv4 DVIPAs, and the only behavior for IPv6 DVIPAs.

VIPADEFINE MOVEABLE WHENIDLE

If an IPv4 DVIPA was originally configured with MOVEABLE WHENIDLE, the following events occur:

- If it appears that there are no active connections to the DVIPA on the backup stack:
  – The DVIPA is removed from the HOME list on the backup stack and reverts to backup status.
  – The restarted stack assumes ownership of the DVIPA by adding it to its HOME list and notifying the routers.
- If there are existing connections to the DVIPA on the backup stack:
  – Ownership of the DVIPA remains with the backup stack. The DVIPA on the restarting stack is placed in backup status at the head of the backup list for the DVIPA.
  – The backup stack periodically checks to see if it has any active connections to the DVIPA.

    When or if it appears that there are no active connections for the DVIPA, the following events occur:

    - The DVIPA is removed from the HOME list on the backup stack and reverts to backup status.
    - The restarted stack assumes ownership of the DVIPA by adding it to its HOME list and notifying the routers.

    Notes:

    1. WHENIDLE is supported only for IPv4 DVIPAs.
    2. A small period of time exists between the check for connections and the movement of the dynamic VIPA to the restarted stack. If connections are made to the old host (the backup stack) in this interval, they will be broken.
    3. During the time that TCP/IP is periodically checking for connections, TCP/IP does not refuse new connections because this would be the same as an outage. If moving the work back to the restarted stack is more important than maintaining uninterrupted service to all clients, then the system operator can use the VARY TCPIP,,OBEYFILE command to delete the dynamic VIPA on the backup stack with the VIPADELETE profile statement. This causes the restarted stack to immediately activate the DVIPA. Optionally, the data set specified on the command can contain a VIPABACKUP statement following the VIPADELETE statement. This will restore the stack as a backup stack.

Movement of unique application-instance (BIND)

A dynamic VIPA is created when any application binds to a nonexistent, specific IP address falling within a configured VIPARANGE on that stack.

In the case of a stack failure, the same application could be started on another stack and (assuming the new stack also has an appropriate VIPARANGE configured) when the application binds to the same IP address, the dynamic VIPA is created on the second stack. Future client connections to that IP address are routed to the second stack where the application is now running.

However, if the same (or a different) application is started on a second stack and attempts to create the same dynamic VIPA using a bind() while it exists on the first stack, the end result is determined by how the VIPARANGE was configured on the stack where the first bind() occurred.

VIPARANGE (DEFINE) MOVEABLE NONDISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE NONDISRUPTIVE, the following events occur:

- If the EZB.BINDDVIPARANGE.*sysname.tcpname* security profile is defined, or if the SAF parameter is specified on the VIPARANGE statement and the EZB.BINDDVIPARANGE.*sysname.tcpname.resname* security profile is defined, the application issuing the bind() call must have READ access to the appropriate resource. If the SAF parameter is specified but the corresponding resource profile is not defined, then the bind fails. If the security profile is not defined but the same VIPA range is configured on another stack, any application on that stack can cause the DVIPA to move there by issuing a bind() call to that DVIPA.
- The DVIPA ownership is immediately transferred to the second stack which adds the DVIPA to its HOME list and dynamically notifies the routers. This stack will now receive all new connections for the DVIPA.
- At the same time, the first stack notifies the routers that it no longer is the owner of the DVIPA, and puts the DVIPA into `moving` status. The DVIPA remains in `moving` status (and in the first stack's HOME list) until the application closes the socket.
- Existing connections on the first stack are preserved. If the second stack receives packets intended for existing connections, it routes the packets to the first stack.

Notes:

1. To ensure preservation of existing connections on the prior owning stack, you must define DYNAMICXCF on both stacks, on the IPCONFIG statement for IPv4 dynamic VIPAs or on the IPCONFIG6 statement for IPv6 dynamic VIPAs.
2. NONDISRUPTIVE is the default for V2R10 and later, and is the only option supported for IPv6.
3. Both stacks must be running V2R10 or later to get non-disruptive behavior. If either stack is running V2R8, the result will be as described in “VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE ” on page 401.

VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE DISRUPTIVE (or if either stack is running V2R8), the following events occur:

- The bind() request for the application on the second stack will fail.
- The DVIPA on the first stack is not affected.

Notes:

1. If movement of the application from the first to the second stack is intended, the application must be ended on the first stack before it is started on the second stack.
2. DISRUPTIVE is supported only for IPv4 DVIPAs.

Defining a security profile for binding to DVIPAs in the VIPARANGE statement

You can define a System Authorization Facility (SAF) resource profile in the SERVAUTH class to control access to dynamic VIPAs (DVIPAs) in a VIPARANGE statement in the following ways:

- Control which applications can bind to create a DVIPA
- Control whether an application can bind to create a DVIPA within a specific VIPARANGE subnet

Rule: If you use the BIND parameter on the PORT statement to create a DVIPA within a VIPARANGE subnet, and you include the SAF parameter on the PORT statement, creation of the DVIPA is controlled by both the SAF resource profile associated with the VIPARANGE statement and the SAF resource profile associated with the PORT statement. For more information about the SAF parameter on the PORT statement, see z/OS Communications Server: IP Configuration Reference.

For information about controlling which applications can issue a SIOCSVIPA ioctl call, a SIOCSVIPA6 ioctl call, or call the MODDVIPA utility to create a DVIPA, see "Defining a security profile for SIOCSVIPA, SIOCSVIPA6, and MODDVIPA" on page 377.

This information includes examples that use RACF. If you are using another security product, see the documentation for that product for the equivalent commands. For more information about RACF, see z/OS Security Server RACF Security Administrator's Guide.

Steps for controlling which applications can bind to create a DVIPA

You can define a System Authorization Facility (SAF) resource profile in the SERVAUTH class to control which applications can bind to create dynamic VIPAs (DVIPAs) in a VIPARANGE statement.

Procedure

Perform the following steps to control whether an application can bind to create a DVIPA:

1. Define the EZB.BINDDVIPARANGE.*sysname*.*tcpname* resource profile in the SERVAUTH class.

   For example, you can define the resource profile using the following RACF command:

   ```
   RDEFINE SERVAUTH (EZB.BINDDVIPARANGE.sysname.tcpname) UACC(NONE)
   ```

2. Give the user ID that is associated with the application READ access to the resource by issuing the following RACF command:

   ```
   PERMIT EZB.BINDDVIPARANGE.sysname.tcpname  ACCESS(READ) CLASS(SERVAUTH) ID(userid)
   ```

3. Refresh the profile by issuing the following RACF command:

   ```
   SETROPTS RACLIST(SERVAUTH) REFRESH
   ```

Results

In this example, *sysname* is the name of the MVS system, *userid* is the user ID that is associated with the application, and *tcpname* is the job name of the TCP/IP started task.

The job name for started tasks, such as TCP/IP, is derived depending on how it is started:

- If the START command is issued with the name of a member in a cataloged procedure library (for example, S TCPIPX), the job name will be the member name (for example, TCPIPX).
- If the member name on the START command is qualified by a started task identifier (for example, S TCPIPX.ABC), the job name will be the started task identifier (for example, ABC). The started task identifier is not visible to all MVS components, but TCP/IP uses it to build the RACF resource name.
- The JOBNAME parameter can also be used on the START command to identify the job name (for example, S TCPIPX,JOBNAME=XYZ).
- The JOBNAME can also be included on the JOB card.

Results:

- If this resource profile is not defined, then the bind is processed.
- If this resource profile is defined and the user ID has READ access to the resource, then the bind is processed.
- If this resource profile is defined and the user ID does not have READ access to the resource, then the bind fails with a permission denied error, regardless of whether the user is a superuser.

Steps for controlling whether an application can bind to create a DVIPA within a specific VIPARANGE subnet

You can define a System Authorization Facility (SAF) resource profile in the SERVAUTH class to control which applications can bind to create dynamic VIPAs (DVIPAs) within a specific VIPARANGE subnet.

Procedure

Perform the following steps to control whether an application can bind to create a DVIPA within a specific VIPARANGE subnet for use by only that application:

1. Define the EZB.BINDDVIPARANGE.*sysname.tcpname.resname* resource profile in the SERVAUTH class.

   For example, to protect a VIPARANGE subnet for application APPL1, you can define the EZB.BINDDVIPARANGE.*sysname.tcpname*.APPL1 resource profile using the following command:

   ```
   RDEFINE SERVAUTH (EZB.BINDDVIPARANGE.sysname.tcpname.APPL1) UACC(NONE)
   ```

2. On the VIPARANGE statement, include the SAF parameter with a *resname* value that matches the *resname* value of the EZB.BINDDVIPARANGE.*sysname.tcpname.resname* resource profile.

   For example:

   ```
   VIPARANGE DEFINE 255.255.255.255 10.10.10.1 SAF APPL1
   ```

   For more information about the SAF parameter on the VIPARANGE statement, see z/OS Communications Server: IP Configuration Reference.

3. Give the user ID that is associated with the application READ access to the EZB.BINDDVIPARANGE.*sysname.tcpname.resname* resource.

   For example:

   ```
   PERMIT EZB.BINDDVIPARANGE.sysname.tcpname.APPL1 ACCESS(READ) CLASS(SERVAUTH)
   ID(userid)
   ```

4. Refresh the resource profile by issuing the following RACF command:

   ```
   SETROPTS RACLIST(SERVAUTH) REFRESH
   ```

Results

In this example, *sysname* is the name of the MVS system, *userid* is the user ID that is associated with the application, and *tcpname* is the job name of the TCP/IP started task. For information about how the job name for a started task, such as TCP/IP, is determined, see "Steps for controlling which applications can bind to create a DVIPA" on page 402.

Results:

- If the SAF parameter is specified and the user ID has READ access to the resource, then the bind is processed.
- If the SAF parameter is specified and the user ID does not have READ access to the resource, then the bind fails.
- If the SAF parameter is specified but the resource profile is not defined, then the bind fails.
- All of the following profile specifications that include wildcard values match the EZB.BINDDVIPARANGE.*sysname.tcpname.resname* resource profile name:
  - EZB.BINDDVIPARANGE.*.*
  - EZB.BINDDVIPARANGE.**
  - EZB.BINDDVIPARANGE.*.*.*

  The most specific match to a resource profile is always used.

Movement of a unique APF-authorized application instance (ioctl)

APF-authorized applications running under a user ID with superuser authority (or that have access through the MODDVIPA security profile) have the ability to activate a dynamic VIPA with the SIOCSVIPA or SIOCSVIPA6 ioctl command, either within the application itself or by invoking the MODDVIPA utility. Because this is a controlled environment, it is assumed configuration errors are minimized or avoided and the usage is correct. Thus, even if the requested DVIPA is currently active on another TCP/IP stack via BIND() or ioctl(), the DVIPA will be immediately activated on this stack. What happens on the other stack is determined by how the VIPARANGE was configured on that stack.

VIPARANGE (DEFINE) MOVEABLE NONDISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE NONDISRUPTIVE, the following events occur:

- The DVIPA ownership is immediately transferred to the second stack which adds the DVIPA to its HOME list and dynamically notifies the routers.
- At the same time, the first stack notifies the routers that it no longer is the owner of the DVIPA, and puts the DVIPA into `moving` status. The DVIPA remains in `moving` status (and in the first stack's HOME list) until the DVIPA is deleted on that stack with the VIPADELETE profile statement or the SIOCSVIPA or SIOCSVIPA6 ioctl DELETE option.
- Existing connections on the first stack are preserved. If the second stack receives packets intended for existing connections, it will route the packets to the first stack.

Notes:

1. To ensure preservation of existing connections on the prior owning stack, you must define DYNAMICXCF on both stacks, on the IPCONFIG statement for IPv4 dynamic VIPAs or on the IPCONFIG6 statement for IPv6 dynamic VIPAs.
2. NONDISRUPTIVE is the default for V2R10 and later.
3. Both stacks must be running V2R10 or later to get non-disruptive behavior. If either stack is running V2R8, the result will be as described in “VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE” on page 404.

VIPARANGE (DEFINE) MOVEABLE DISRUPTIVE

If the first stack is configured with VIPARANGE MOVEABLE DISRUPTIVE (or if either stack is running V2R8), the following events occur:

- The ioctl request for the application on the second stack succeeds. The DVIPA is added to the HOME list on the second stack, and the routers are dynamically notified.
- The DVIPA on the first stack is deleted.

Notes:

1. Any existing connections to the DVIPA on the first stack are broken.
2. DISRUPTIVE is IPv4 only.

Same dynamic VIPA for VIPADEFINE and BIND(), SIOCSVIPA or SIOCSVIPA6 ioctl, or MODDVIPA utility

Regardless of careful implementation, it is possible that the same IP address is inadvertently selected for VIPADEFINE and for use with BIND(), SIOCSVIPA or SIOCSVIPA6 ioctl, or the MODDVIPA utility. Because the application scenarios are quite different, this must be an error.

If this duplicate DVIPA address conflict occurs on the same TCP/IP, the second attempt might fail. If an IP address is specified in a VIPADEFINE, and that same IP address has already been activated on the TCP/IP by an application using BIND(), the SIOCSVIPA or SIOCSVIPA6 ioctl, or the MODDVIPA utility, the VIPADEFINE will be rejected during VARY TCPIP,,OBEYFILE command processing. If an IP address is activated with VIPADEFINE, and the application does a BIND(), ioctl(), or the MODDVIPA utility is used,

the BIND() will succeed, but the ioctl() will fail with a nonzero errno and the MODDVIPA utility will set a nonzero condition to indicate that the IP address already exists.

The same situation could also occur on two different TCP/IPs in the sysplex. Because the TCP/IPs are exchanging information among themselves, if the two attempts are far enough apart in time, the second attempt will be caught immediately and rejected. However, it is possible that the attempt will be made almost simultaneously on two different TCP/IPs, such that neither TCP/IP is yet aware of the attempt on the other TCP/IP. If both attempt such an activation, and the exchange of information then shows a conflict, the internal sysplex time stamps are used to determine which attempt was really first. The one that appears to be first is allowed to continue, and the dynamic VIPA is deleted from the later TCP/IP. While such a simultaneous attempt is somewhat unpredictable in respect to which one will succeed, the dynamic VIPA will remain active on only one TCP/IP, and examination of messages will indicate which TCP/IP successfully created the DVIPA and on which TCP/IP it was rejected.

Dynamic VIPA creation results

Table 20 on page 405 summarizes the results of attempting to create a dynamic VIPA when it (or the same IP address for HOME statement) already exists in the sysplex.

Table 20. Summary of dynamic VIPA creation results

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
bind()	bind()	Second bind() succeeds, but no new VIPA is created.	If both stacks are running V2R10 or later, and the first BIND DVIPA was created with MOVEABLE NONDISRUPTIVE: • On stack 2, bind() succeeds • On stack 1, the BIND VIPA remains in the HOME list (unadvertised) and any existing connections are preserved • New connections to that IP address go to the application on stack 2. Otherwise, second bind fails.
bind()	ioctl()	ioctl() fails with warning condition code, but the application associated with the ioctl is still able to use the dynamic VIPA.	ioctl() succeeds, bind is deleted (even if BIND DVIPA was created as MOVEABLE NONDISRUPTIVE)
bind()	VIPADEFINE	VIPADEFINE fails.	VIPADEFINE fails.
bind()	VIPABACKUP	VIPABACKUP fails.	VIPABACKUP fails.
bind()	HOME	See note at end of table.	See note at end of table.
ioctl()	bind()	bind() succeeds, no new VIPA is created.	bind() fails.

Table 20. Summary of dynamic VIPA creation results (continued)

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
ioctl()	ioctl()	Second ioctl() fails with warning condition code, but the application associated with the ioctl is still able to use the dynamic VIPA.	Second ioctl() succeeds. If both stacks are running V2R10 or later, and the ioctl DVIPA on stack 1 was created with MOVEABLE NONDISRUPTIVE, the DVIPA on stack 1 remains in the HOME list (unadvertised) and any existing connections are preserved. Otherwise, the ioctl DVIPA on stack 1 is deleted and any existing connections are broken.
ioctl()	VIPADEFINE	VIPADEFINE fails.	VIPADEFINE fails.
ioctl()	VIPABACKUP	VIPABACKUP fails.	VIPABACKUP fails.
ioctl()	HOME	See note at end of table.	See note at end of table.
VIPADEFINE	bind()	bind() succeeds, but no new VIPA is created.	bind() fails.
VIPADEFINE	ioctl()	ioctl() fails.	ioctl() fails.

Table 20. Summary of dynamic VIPA creation results (continued)

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
VIPADEFINE	VIPADEFINE	If the second VIPADEFINE statement is an exact duplicate of the first, the second VIPADEFINE is ignored with no error message. For IPv4, the second VIPADEFINE fails if it has different options or a different mask than the first VIPADEFINE specified. For IPv6, the second VIPADEFINE fails if the interface name is already defined with a different address or the address is already defined for a different interface name.	For IPv4, the second VIPADEFINE succeeds but activation on stack 2 might be deferred. If both stacks are running V2R10 or later, and the DVIPA was created on stack 1 as MOVEABLE IMMEDIATE: • Second VIPADEFINE is activated immediately • Any connections to the DVIPA on stack 1 are preserved. (DVIPA stays in HOME list unadvertised) Otherwise, the second VIPADEFINE activation is deferred until there are no connections on stack 1, at which point, stack 1 reverts to backup status. For IPv6, if both the interface name and address on the second VIPADEFINE match the first VIPADEFINE, the DVIPA is activated on stack 2. Any connections to the DVIPA on stack 1 are preserved. If either the interface name or address is the same but the other is not, the second VIPADEFINE fails.
VIPADEFINE	VIPABACKUP	VIPABACKUP fails.	Both succeed.
VIPADEFINE	HOME	See note at end of table.	See note at end of table.
VIPABACKUP	bind()	bind() fails.	If the IP address is already active on the bind() stack, the bind() will succeed. Otherwise, the bind() fails.
VIPABACKUP	ioctl()	ioctl() fails.	ioctl() fails.
VIPABACKUP is backup status	VIPADEFINE	VIPADEFINE succeeds, replaces the VIPABACKUP. For IPv6, both the interface name and the address on the VIPADEFINE must match the VIPABACKUP. If one matches and the other does not match, the VIPADEFINE fails.	VIPADEFINE succeeds. For IPv6, both the interface name and the address on the VIPADEFINE must match the VIPABACKUP. If one matches and the other does not match, the VIPADEFINE fails.

Table 20. Summary of dynamic VIPA creation results (continued)

First action	Second action	Result if second action is on the same stack	Result if the second action is on a different stack within the sysplex
VIPABACKUP in active status (after takeover)	VIPADEFINE	VIPADEFINE rejected	For IPv4, the VIPABACKUP DVIPA is MOVEABLE IMMEDIATE or WHENIDLE depending how the original VIPADEFINE DVIPA was created. If both stacks are running V2R10 or later, and the VIPABACKUP DVIPA is MOVEABLE IMMEDIATE: • The VIPADEFINE is activated immediately. • Any connections to the DVIPA on stack 1 are preserved (DVIPA stays in HOME list unadvertised). • When there are no more connections, stack 1 reverts to backup status. Otherwise, the VIPADEFINE activation on stack 2 is deferred until there are no active connections on stack 1, at which point stack 1 reverts to backup status. For IPv6, if both the interface name and the address on the VIPADEFINE match the VIPABACKUP, the DVIPA is activated on stack 2. Any connections to stack 1 are preserved. If only one matches, the VIPADEFINE fails.
VIPABACKUP	VIPABACKUP	If the DVIPA is in backup status on this stack, the second VIPABACKUP succeeds. If the DVIPA is in active status on this stack (for example, after a takeover), a VIPABACKUP with a different rank will be rejected. For IPv6, both the interface name and address must match the first VIPABACKUP. If one matches and the other is different, the second VIPABACKUP fails.	Second VIPABACKUP succeeds. For IPv6, both the interface name and address must match the first VIPABACKUP. If one matches and the other is different, the second VIPABACKUP fails.
VIPABACKUP	HOME	See note at end of table.	See note at end of table.

Table 20. Summary of dynamic VIPA creation results (continued)			
First action	**Second action**	**Result if second action is on the same stack**	**Result if the second action is on a different stack within the sysplex**
HOME	bind()	bind() succeeds, but no new VIPA is created.	bind() fails.
HOME	ioctl()	ioctl() fails.	ioctl() fails.
HOME	VIPADEFINE	VIPADEFINE fails.	VIPADEFINE fails.
HOME	VIPABACKUP	VIPABACKUP fails.	VIPABACKUP fails.
Note: Defining the same IP address in the HOME list as an existing dynamic VIPA will not be rejected by the TCP/IP stack, but it is likely to cause routing problems.			

TIER1, TIER2, and CPCSCOPE keyword DVIPA contention resolution

Sysplex distributor supports load balancing across z/OS images in a sysplex, as described in "Sysplex distribution optimizations for multi-tier z/OS workloads" on page 487; the TIER1, TIER2, and CPCSCOPE keywords are used with this support. Use the following guidelines to define dynamic VIPAs (DVIPAs) that use these keywords.

Guidelines:

- If you want to change the TIER1, TIER2, or CPCSCOPE definition for a DVIPA, delete the existing definition of the DVIPA from the sysplex first; the keyword usage for a DVIPA must be the same throughout the sysplex. For example:
  - If there is an existing VIPADEFINE or VIPABACKUP statement for DVIPA1 using the TIER2 keyword, use a VIPADELETE statement to remove this DVIPA before using a VIPADEFINE TIER1 statement for DVIPA1.
  - If there is an existing VIPADEFINE or VIPABACKUP statement for DVIPA1 that does not use the TIER1, TIER2, or CPCSCOPE keywords, use a VIPADELETE statement to remove this DVIPA before defining DVIPA1 with one of these keywords.
- When you are defining a DVIPA with the CPCSCOPE keyword, and the DVIPA exists in a different central processor complex (CPC), delete it from that CPC first; a DVIPA with the CPCSCOPE keyword can exist only on stacks in the same CPC. For example, if DVIPA1 was defined using a VIPABACKUP CPCSCOPE statement on a stack in CPC1, use a VIPADELETE statement to remove this DVIPA before using a VIPABACKUP CPCSCOPE statement for a stack in CPC2.

If you do not follow these guidelines, the changes to DVIPA creation results that occur when the TIER1, TIER2, or CPCSCOPE keyword is changed, added, or removed from a DVIPA that already exists in the sysplex are described in Table 21 on page 410, Table 22 on page 411, and Table 23 on page 412.

If you create a new definition for an existing DVIPA on the same stack, either in the same profile or later using the VARY TCPIP,,OBEYFILE command, see Table 21 on page 410. When the new definition is processed, it cannot change the TIER1, TIER2, or CPCSCOPE usage; if this is attempted, the new definition is rejected.

Table 21. DVIPA contention resolution for DVIPA definitions on the same stack		
Original definition	**Later definition**	**Result**
VIPADEFINE or VIPABACKUP DVIPA1 TIER1	VIPADEFINE or VIPABACKUP DVIPA1 TIER1	Later definition replaces the original definition
VIPADEFINE or VIPABACKUP DVIPA1 TIER2	VIPADEFINE or VIPABACKUP DVIPA1 TIER2	
VIPADEFINE or VIPABACKUP DVIPA1 TIER2 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 TIER2 CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 TIER1	VIPADEFINE or VIPABACKUP DVIPA1 TIER2 or CPCSCOPE	Configuration mismatch; later definitions are rejected
VIPADEFINE or VIPABACKUP DVIPA1 TIER2	VIPADEFINE or VIPABACKUP DVIPA1 TIER1 or CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 TIER1 or TIER2	
VIPADEFINE or VIPABACKUP DVIPA1	VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1	

When two stacks define the same DVIPA, contention resolution is as follows:

- If there is a TIER1, TIER2, or CPCSCOPE mismatch for the DVIPA, the stack with the later timestamp cannot change the keyword usage, and either rejects its own definition during profile processing or deletes its definition when it learns of the DVIPA definition with the earlier timestamp.
- If there is a mismatch between a DVIPA definition that uses a TIER1, TIER2, or CPCSCOPE keyword and a DVIPA definition that does not use any of these keywords, possible outcomes are as follows:
  - If the second stack has already received the mismatched DVIPA definition from the first stack prior to profile processing, contention resolution depends on the release level of stack 2:
    - If stack 2 is V1R11 or later, the earlier timestamp wins and stack 2 rejects all DVIPA1 definitions (including the VIPADISTRIBUTE statement)
    - If stack 2 is V1R10 or earlier, it is not aware of the new keywords and the mismatch and it defines the DVIPA; stack 1 deletes its definition when it learns that stack 2 defined the DVIPA
  - After two stacks have successfully defined DVIPAs with mismatching definitions, the stack that uses the keywords deletes all of its definitions for that DVIPA, because the stack that uses no keywords might be using V1R10 or earlier.

Table 22 on page 411 summarizes DVIPA contention resolution between stacks in the same CPC.

Table 22. DVIPA contention resolution between stacks in the same CPC

Stack 1 — Earlier timestamp	**Stack 2 — Later timestamp**	**Result**
VIPADEFINE DVIPA1 TIER1	VIPADEFINE DVIPA1 TIER1	Later timestamp wins and stack 1 becomes a backup for DVIPA1
VIPADEFINE DVIPA1 TIER2	VIPADEFINE DVIPA1 TIER2	
VIPADEFINE DVIPA1 TIER2 CPCSCOPE	VIPADEFINE DVIPA1 TIER2 CPCSCOPE	
VIPADEFINE DVIPA1 CPCSCOPE	VIPADEFINE DVIPA1 CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 TIER1	VIPADEFINE or VIPABACKUP DVIPA1 TIER2 or CPCSCOPE	Configuration mismatch; earlier timestamp wins and stack 2 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement)
VIPADEFINE or VIPABACKUP DVIPA1 TIER2	VIPADEFINE or VIPABACKUP DVIPA1 TIER1 or CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 TIER1 or TIER2	
VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1	Configuration mismatch during profile processing for stack 2: • If stack 2 is V1R11 or later, the earlier timestamp wins and stack 2 rejects all DVIPA1 definitions (including the VIPADISTRIBUTE statement) • If stack 2 is V1R10 or earlier, it is not aware of the new keywords and the mismatch and it defines the DVIPA; stack 1 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement) when it learns that stack 2 defined the DVIPA

Table 22. DVIPA contention resolution between stacks in the same CPC (continued)		
Stack 1 — Earlier timestamp	**Stack 2 — Later timestamp**	**Result**
VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1	Configuration mismatch after profile processing for stack 2 is completed; definition with no keywords wins and stack 1 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement)
VIPADEFINE or VIPABACKUP DVIPA1	VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	Configuration mismatch after profile processing for stack 2 is completed; definition with no keywords wins and stack 2 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement)

Table 23 on page 412 summarizes additional contention resolution rules used when the stacks are in different CPCs.

Table 23. DVIPA contention resolution between stacks in different CPCs		
Stack 1 CPC 1 — Earlier timestamp	**Stack 2 CPC 2 — Later timestamp**	**Result**
VIPADEFINE DVIPA1 TIER1	VIPADEFINE DVIPA1 TIER1	Later timestamp wins and stack 1 becomes a backup for DVIPA1
VIPADEFINE DVIPA1 TIER2	VIPADEFINE DVIPA1 TIER2	
VIPADEFINE or VIPABACKUP DVIPA1 TIER2 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 TIER2 CPCSCOPE	Earliest Timestamp wins and stack 2 deletes its VIPADEFINE or VIPABACKUP statement because the DVIPA can have VIPADEFINE and VIPABACKUP definitions only on stacks in the same CPC
VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	

Table 23. DVIPA contention resolution between stacks in different CPCs (continued)

Stack 1 CPC 1 — Earlier timestamp	Stack 2 CPC 2 — Later timestamp	Result
VIPADEFINE or VIPABACKUP DVIPA1 TIER1	VIPADEFINE or VIPABACKUP DVIPA1 TIER2 or CPCSCOPE	Configuration mismatch; earlier timestamp wins and stack 2 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement)
VIPADEFINE or VIPABACKUP DVIPA1 TIER2	VIPADEFINE or VIPABACKUP DVIPA1 TIER1 or CPCSCOPE	
VIPADEFINE or VIPABACKUP DVIPA1 CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1 TIER1 or TIER2	
VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1	Configuration mismatch during profile processing for stack 2: • If stack 2 is V1R11 or later, the earlier timestamp wins and stack 2 rejects all DVIPA1 definitions (including the VIPADISTRIBUTE statement) • If stack 2 is V1R10 or earlier, it is not aware of the new keywords and the mismatch and it defines the DVIPA; stack 1 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement) when it learns that stack 2 defined the DVIPA
VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	VIPADEFINE or VIPABACKUP DVIPA1	Configuration mismatch after profile processing for stack 2 is completed; definition with no keywords wins and stack 1 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement)

Table 23. DVIPA contention resolution between stacks in different CPCs (continued)		
Stack 1 CPC 1 — Earlier timestamp	**Stack 2 CPC 2 — Later timestamp**	**Result**
VIPADEFINE or VIPABACKUP DVIPA1	VIPADEFINE or VIPABACKUP DVIPA1 TIER1, TIER2, or CPCSCOPE	Configuration mismatch after profile processing for stack 2 is completed; definition with no keywords wins and stack 2 deletes all DVIPA1 definitions (including the VIPADISTRIBUTE statement)

IPv6 considerations

This topic discusses special considerations for IPv6.

VIPARANGE

The MOVeable DISTRUPTive parameter is not supported for IPv6. To prevent an unauthorized application from creating a dynamic VIPA with a bind() call, you can add the EZB.BINDDVIPARANGE.*sysname.tcpname* resource profile, or you can use the SAF parameter along with the EZB.BINDDVIPARANGE.*sysname.tcpname.resname* resource profile. If the security check fails, the bind() request fails, and if the security check is successful, the bind() request is accepted.

VIPADEFINE and VIPABACKUP

VIPABACKUP must specify an interface name and IPv6 address. The interface name and IPv6 address pair must match any preexisting dynamic VIPA definitions for either that interface name or IPv6 address. If using VIPABACKUP for an initial DVIPA activation, neither the interface name nor the IPv6 address can preexist. Furthermore, for the takeover and backup functions to succeed, if a corresponding VIPADEFINE is to be subsequently activated, the interface name and IPv6 address on the VIPADEFINE statement must match the VIPABACKUP statement.

Unique application-instance scenario and IPv6-enabled applications

A single instance of an IPv6-enabled TCP application can typically handle communications with IPv4 or IPv6 partner applications. IPv6-enabled TCP server applications usually accomplish these communications by using a single AF_INET6 socket to accept connections from both IPv4 and IPv6 clients. As a result, IPv6-enabled applications have some additional configuration considerations when compared to IPv4-only applications, as they require both an IPv4 DVIPA and an IPv6 DVIPA to be configured. If the application is using the MODDVIPA utility or the SIOCSVIPA IOCTL to define the DVIPA, it must be modified as follows:

- If you are using the MODDVIPA utility, an additional set of invocations for this utility is needed to define and delete the IPv6 DVIPA to be associated with the application.
- If the application is using the SIOCSVIPA IOCTL, it must be modified to also issue the SIOCSVIPA6 IOCTL to define and delete the IPv6 DVIPA to be associated with the application.

Applications that are currently relying on defining the DVIPA dynamically by using the bind() socket API or the BIND parameter on the PORT reservation statement have some additional considerations, as a single socket cannot be bound to more than one IP address.

The following options can help alleviate this issue:

- Instead of relying on the bind to trigger the DVIPA definition, the application or configuration can be modified to allow the application to bind to the IPv6 unspecified address (in6addr_any). The MODDVIPA

utility can then be used to define and delete both the IPv4 and IPv6 DVIPAs associated with the application.

Another similar alternative is for the application to issue the SIOCSVIPA or SIOCSVIPA6 IOCTL for both of these DVIPAs. If two or more application instances are to bind their listeners to the same port and the IPv6 unspecified address, each DVIPA can be created with affinity to the application instance that created it. Such affinity ensures that a connection request is sent to the correct listener. A connection request for a DVIPA that is created with affinity is sent to a TCP Listener if the application instance that created the DVIPA issues the bind() call. If no matching listener is available, normal shareport load balancing is used to choose the best available listener. See "Configuring the unique application-instance scenario" on page 373 for information about creating a DVIPA with affinity.

- Start two instances of the application, one binding to the IPv4 DVIPA and the other binding to the IPv6 DVIPA, assuming that multiple instances of the same application can be started within the same MVS image or across different images in the sysplex. This configuration allows IPv4 clients to reach the application instance that is bound to the IPv4 DVIPA and IPv6 clients to reach the application instance that is bound to the IPv6 DVIPA.
- In some cases, it might be possible to modify the application so that it can create two sockets, one for IPv6 communications and one for IPv4 communications. Each socket can then be bound to the IPv4 or IPv6 DVIPA as appropriate, and a single instance of the application can handle both IPv4 and IPv6 clients. This solution might be less desirable because it also involves application code changes.

VIPAs, OSA-Express QDIO, and Spanning Tree Protocol

Spanning Tree Protocol (STP) can potentially impact environments where both OSA-Express QDIO and VIPA are deployed.

With dynamic VIPA, a TCP/IP address takeover occurs when a TCP/IP stack fails, or a static or dynamic VIPA is manually moved by operator intervention. In either case, when a static or dynamic VIPA moves, the IP address and respective workloads are taken over by another TCP/IP stack through other OSA-Express Ethernet devices running on a different server. When the original TCP/IP stack and respective OSA-Express QDIO devices are returned to operation, both the IP address and respective workload traffic are taken back by the recovered TCP/IP stack.

If the network bridge or switch is not configured properly, packets can get lost in the network or be blocked by the networking equipment. This is a result of a physical looping condition identified by the STP, or expired OSA-Express QDIO timers due to the increased latencies associated with blocked ports or delayed packets. In these cases, static or dynamic VIPA connectivity can fail.

When configuring STP, use care in the bridge or switch configuration to avoid or minimize potential loop conditions. For example, if the STP is not an integral component of the overall network, disabling the STP on all of the Virtual LANs (VLANs) that connect to OSA-Express QDIO devices will help avoid the problem.

Also, some networking switches provide a mechanism for suppressing the STP's listening and learning states on specific switch ports. For example, Cisco Systems provides an STP configuration feature called PortFast that can place specific switch ports into forwarding state as soon as a link is detected. Without PortFast enabled, a switch port has to transition through the listening and learning stages (30 seconds total) of Spanning Tree reconvergence before the switch port can actually pass valid traffic. For PortFast capable Cisco switches, enabling PortFast on all switch ports that OSA-Express QDIO is connected into allows network administrators to both preserve their original Spanning Tree configuration by not having to change it, while also providing a viable mechanism to avoid potential static or dynamic VIPA problems.

For more information on configuring STP and immediate port forwarding, see the bridge or switch operational manual.

Mixture of types of dynamic VIPAs within subnets

Any particular IP address can be used in only one way as a dynamic VIPA. A dynamic VIPA can be defined either with VIPADEFINE or by application action within a valid VIPARANGE, but not both. However, within a subnet defined as a VIPARANGE, some IP addresses can be used for VIPADEFINE, and others may be

assigned to unique application instances, without conflict, as long as the limit of a total of 1024 active and backup dynamic VIPAs on a single TCP/IP is not exceeded. TCP/IP will make no attempt to reject a VIPADEFINE dynamic VIPA that also falls within a VIPARANGE. This allows installations with limited availability of IP addresses to assign individual addresses to either application scenario, without having to define separate subnets and use up additional IP addresses in that manner.

MVS failure and sysplex failure management

The TCP/IPs in a sysplex use MVS XCF messaging to exchange information about dynamic VIPAs. When a TCP/IP fails or is ended by operator command, but the underlying MVS remains active, the other TCP/IPs are immediately notified, and takeover of VIPADEFINE dynamic VIPAs is automated and very fast. In addition, each TCP/IP stack monitors for local problem conditions that might cause it to leave the TCP/IP sysplex group, at which point the other TCP/IPs will be immediately notified. For more information, see "Sysplex problem detection and recovery" on page 452.

However, when an MVS fails, there is normally an operator message on the console requiring a response (WTOR). Until this response is made by an operator or automation, the other MVS systems do not notify the remaining TCP/IPs in the sysplex of the failure of the TCP/IP on the failing MVS. This can delay automated backup of dynamic VIPAs defined with VIPADEFINE. Sysplex Failure Management (SFM) can be used to automate the required response to the console message of the failing MVS. See z/OS MVS Setting Up a Sysplex for information on how to set up SFM to avoid the requirement for a manual response and speed backup of dynamic VIPAs defined with VIPADEFINE.

For more information, see z/OS Communications Server: IP Diagnosis Guide.

Applications and dynamic VIPAs

While most applications support multiple instances in a sysplex, very few applications expect IP addresses to move around under them. TCP applications use TCP connections to form a relationship between particular client and server instances to exchange data over an extended period. They rely on notification of TCP connection termination to initiate recovery and to reestablish a new relationship (possibly from a client to a different server). Conversely, most UDP applications do the equivalent function at the application layer. Movement of an IP address to a different server could be confusing to the client, unless the new server also is aware of the state of the client work.

UDP applications whose interactions consist of atomic interactions (a single request followed by one or more responses, with no state information maintained at the server between requests) can use dynamic VIPAs in the multiple application-instance scenario. However, if the server application maintains state information between interactions (for example, NFS), then moving a dynamic VIPA to another server might not work unless the client/server application protocol can detect the discontinuity. In that case, the unique application-instance scenario might apply, which would require the restart of the server instance on another TCP/IP.

In addition, the following types of work are not appropriate for distribution with distributed dynamic VIPA:

- Applications that establish affinity with a particular TCP/IP stack, such as SNMP.
- FTP servers that receive the PASV or EPSV command for a distributed DVIPA that did not specify SYSPLEXPORTS. The PASV and EPSV commands are supported when SYSPLEXPORTS was specified on the VIPADISTRIBUTE statement of the distributed DVIPA that is the destination IP address being used by the FTP server. These commands request that the FTP server bind() on a data port that is not the default data port, or the one specified on the VIPADISTRIBUTE statement, and to wait for a connection rather than initiate one on receipt of a transfer command (for example, RETR). Because SYSPLEXPORTS cannot be specified for a non-distributed dynamic VIPA, passive mode FTP connections made to a non-distributed dynamic VIPA cannot be recovered when the VIPA is moved through a planned takeover. The control connection remains on the original stack, but attempts to create new data connections for control connections that existed prior to the move will fail. When SYSPLEXPORTS is used with a distributed dynamic VIPA, new data connections are always sent to the same stack containing their control connection. For more information on using the SYSPLEXPORTS parameter in the VIPADYNAMIC block, see z/OS Communications Server: IP Configuration Reference.

- FTP servers that accept connections on an IPv6 distributed DVIPA that does not have SYSPLEXPORTS specified. The FTP server expects an EPSV command for data transfers to or from an IPv6 address, and use of the EPSV command is supported only when SYSPLEXPORTS was specified on the VIPADISTRIBUTE statement of the distributed DVIPA that is the destination IP address being used by the FTP server.

Configuring VIPAs for activation with VIPABACKUP

When adding a dynamic VIPA to a functioning sysplex, the normal order is to add VIPABACKUP statements to backup stacks, and then add a VIPADEFINE statement to the TCP/IP stack where the VIPA should normally be, at which time the VIPA would be activated and ready for use. A VIPADISTRIBUTE statement could be added at the same time. The VIPA would be activated only when the VIPADEFINE was processed on the TCP/IP that would normally own it, and not when a VIPABACKUP is processed on another stack. This is, in part, because parameters on the VIPADEFINE that are needed to activate the VIPA were not found on the VIPABACKUP statement.

However, on the rare occasion that a disaster occurs, it might be necessary to IPL all of the systems in a sysplex. Assuming that many dynamic VIPAs are in use and the VIPADEFINE statements are throughout the available TCP/IP stacks in the sysplex, most of the dynamic VIPAs have a lengthy wait before the owning operating system, TCP/IP, and application are started and fully operational. The intent of dynamic VIPAs defined with VIPADEFINE and VIPABACKUP is to move the dynamic VIPAs under a functioning application as soon as possible. Therefore, optional parameters have been added to the VIPABACKUP statement to allow the dynamic VIPA to be activated when the VIPABACKUP is processed at TCP/IP initialization or VARY TCPIP,,OBEYFILE command processing. If the MOVEABLE IMMEDIATE or MOVEABLE WHENIDLE parameter is specified on a VIPABACKUP profile statement, along with required address mask and optional SERVICEMGR parameters for an IPv4 DVIPA, the dynamic VIPA will be activated when the profile statement is processed, if it is not active elsewhere in the sysplex already. The following notes apply to this case:

- If the MOVEABLE parameter is specified, the address mask must also be specified. The address mask and SERVICEMGR parameters can be specified on a VIPABACKUP statement only if the MOVEABLE parameter has also been specified.
- The first started VIPABACKUP TCP/IP stack that is configured to start a particular dynamic VIPA with VIPABACKUP will activate the dynamic VIPA and own it until the VIPADEFINE stack is started. As long as the first activating VIPABACKUP stack remains operational, starting another stack with a VIPABACKUP statement will not cause the dynamic VIPA to move, even if the second stack has a higher backup rank. Only the initialization of the VIPADEFINE stack will cause the VIPA to move automatically.
- When the VIPADEFINE is eventually processed, its parameters will override the values specified on the VIPABACKUP that activated the dynamic VIPA, if the VIPADEFINE has parameters that are different from those specified on the VIPABACKUP.

It is always a good idea, when able to specify a parameter on both primary and backup, to specify exactly the same values for each. When all the parameters common to VIPADEFINE and VIPABACKUP are specified the same on both for a particular DVIPA, the behavior exhibited when the VIPADEFINE stack comes up will be the same as when the DVIPA is activated first by a VIPABACKUP stack. However, for IPv4, there are several parameters that could be specified inconsistently on the VIPABACKUP and VIPADEFINE statements. If they are specified inconsistently, the following implications are some of the implications by parameter:

- MOVEABLE: If there is any doubt, this parameter should be coded as MOVEABLE IMMEDIATE on the VIPABACKUP stack, so that the stack will maintain connection information to be sent to the VIPADEFINE stack when it comes up. Subsequent operation in the sysplex will be determined by the coding (or defaulting) of MOVEABLE on the VIPADEFINE statement, regardless of what was coded on the VIPABACKUP of the stack that first activated the DVIPA. This is true even if that same stack later takes over the DVIPA from the VIPADEFINE or another stack. (This is the reason MOVEABLE was chosen as the way to activate the new function on VIPABACKUP stacks, rather than a new keyword, to force consideration of the setting of the MOVEABLE parameter for VIPABACKUP stacks.)

- address_mask: The address_mask is used to build BSDROUTINGPARMS statements. If the attached routing daemon is OMPROUTE and there is no corresponding interface definition with the subnet mask parameter in OMPROUTE configuration, OMPROUTE might send a different address mask from the VIPADEFINE stack than it will for the VIPABACKUP stack, which might confuse the routing network. The address_mask should be the same on VIPABACKUP and VIPADEFINE statements for a particular DVIPA.

  Result: Incorrect OMPROUTE configuration, such as missing interface definitions for RIPv1, RIPv2, or OSPF, can lead to unexpected results, especially if OMPROUTE uses defaults for the address mask.

Assuming that dynamic VIPAs are throughout the sysplex, and that critical applications are as well, the first stack to be activated might activate most, if not all, DVIPAs, and might therefore assume a considerable amount of the workload. For any particular application, all of the workload is directed to the first stack supporting that application to be activated, regardless of what other workload will be executing on that same stack or LPAR.

Guideline: Consider the planned sysplex TCP/IP activation sequence, and have only those DVIPAs for critical applications activating on TCP/IP stacks early in the sequence, so that less important work does not interfere with the business-critical applications during sysplex startups.

Example of configuring dynamic and distributed VIPAs

The TCP/IP profiles needed to implement dynamic VIPA (DVIPA) on multiple systems in a sysplex are shown in the following examples. The VIPADEFINE and VIPABACKUP statements allow automatic dynamic VIPA takeover to occur if needed (see "Configuring the multiple application-instance scenario" on page 373), and the VIPARANGE statements allow dynamic VIPAs to be dynamically created by an application or by the MODDVIPA utility (see "Configuring the unique application-instance scenario" on page 373). The VIPADISTRIBUTE statements allow a single VIPA to be shared among several TCP/IP stacks. Including the SOURCEVIPA and TCPSTACKSOURCEVIPA parameters on the IPCONFIG and IPCONFIG6 statements, on each target stack with the same dynamic VIPA specified, enables a single DVIPA address to be used as a sysplex-wide source DVIPA address for outbound TCP connections. The following examples show both IPv4 and IPv6 DVIPAs, and the output is shown in the IPv6-enabled, or long, format.

```
TCPCS
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.1 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0001 INTFID 6:7:8:9
 SOURCEVIPAINT SVIPA1 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

VIPADYNAMIC
  VIPADEFINE 255.255.255.240 201.2.10.11 201.2.10.12
  VIPADEFINE 255.255.255.240 201.2.10.14 201.2.10.15
  VIPADEFINE 255.255.255.240 201.2.10.23
  VIPADISTRIBUTE SYSPLEXPORTS DISTM SERVERWLM 201.2.10.11
   PORT 20 21 DESTIP ALL
  VIPADISTRIBUTE 201.2.10.12 PORT 20 21 DESTIP 193.9.200.2
  VIPADISTRIBUTE DISTMETHOD ROUNDROBIN 201.2.10.14 DESTIP 193.9.200.2
  VIPADISTRIBUTE DISTM WEIGHTEDActive 201.2.10.15 PORT 5000
   DESTIP 193.9.200.2 WEIGHT 10
          193.9.200.3 WEIGHT 20
  VIPADISTRIBUTE TIMEDAFF 30 201.2.10.15 PORT 23 DESTIP ALL
   PROCTYPE CP 20 ZAAP 80 ZIIP 0
  VIPADISTRIBUTE 201.2.10.23 PORT 4000 DESTIP ALL
  VIPABACKUP 100 201.2.10.13
  VIPABACKUP 80 201.2.10.21 201.2.10.22
  VIPARANGE DEFINE 255.255.255.192 201.2.10.192
  VIPADEFINE DVIPA1 2001:0DB8:1::1
  VIPADISTRIBUTE SYSPLEXPORTS DISTMETHOD SERVERWLM DVIPA1 DESTIP ALL
  VIPADEFINE DVIPA2 2001:0DB8:2::2
  VIPADISTRIBUTE TIMEDAFF 45 DISTM ROUNDROBIN DVIPA2 PORT 23 DESTIP ALL
  VIPARANGE VRANGE1 2001:0DB8:3::1/100
ENDVIPADYNAMIC
```

```
TCPCS2
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.2 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0002
```

```
 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

VIPADYNAMIC
  VIPADEFINE 255.255.255.192 201.2.10.13
  VIPABACKUP 100 201.2.10.11 201.2.10.21
  VIPABACKUP 75 201.2.10.12 201.2.10.22
  VIPARANGE DEFINE 255.255.255.192 201.2.10.192
  VIPADEFINE DVIPA3 2001:0DB8:3::3
  VIPABACKUP 200 DVIPA2
  VIPABACKUP 220 DVIPA1
  VIPADISTRIBUTE SYSPLEXPORTS DVIPA1 PORT 23 DESTIP ALL
  VIPABACKUP 10  DVIPA4
ENDVIPADYNAMIC
```

```
TCPCS3
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.3 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0003
 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

VIPADYNAMIC
  VIPADEFINE MOVE IMMED 255.255.255.192 201.2.10.21 201.2.10.22
  VIPABACKUP 10 201.2.10.11 201.2.10.12 201.2.10.13
  VIPARANGE DEFINE 255.255.255.192 201.2.10.192
  VIPADEFINE DVIPA4 2001:0DB8:4::4
  VIPABACKUP 110 DVIPA2
  VIPABACKUP 100 DVIPA1
ENDVIPADYNAMIC
```

```
TCPCS6
IPCONFIG SYSPLEXROUTING SOURCEVIPA TCPSTACKSOURCEVIPA 201.2.10.11
 DYNAMICXCF 193.9.200.6 255.255.255.240 1
IPCONFIG6 DYNAMICXCF 2001:0DB8::151:0006
 SOURCEVIPA TCPSTACKSOURCEVIPA DVIPA1

TCPCS6 does not have dynamic VIPAs defined so it does not
contain a VIPADYNAMIC definition.  It has DYNAMICXCF specified
for IPv4 and IPv6 to enable XCF dynamic support and to allow
TCPCS6 to be a target for dynamic VIPA distribution.
```

Start TCP/IP on each system as shown.

- On system1, start TCPCS and TCPCS2.
- On system2, start TCPCS3, on system3 start TCPCS6.
- On system1, run the MODDVIPA utility to define the DVIPA 201.2.10.193.

```
//TCPDVP   PROC
//*
//*
//TCPDVP  EXEC PGM=MODDVIPA ,REGION=0K,TIME=1440,                       x
//             PARM='POSIX(ON) ALL31(ON)/-p TCPCS -c 201.2.10.193'
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
//SYSERR   DD SYSOUT=*
//SYSERROR DD SYSOUT=*
//SYSDEBUG DD SYSOUT=*
//SYSUDUMP DD SYSOUT=A
//SYSABEND DD SYSOUT=*
//*
//*Run program here
//*
//TCPDVP  EXEC PGM=MODDVIPA ,REGION=0K,TIME=1440,                       x
//             PARM='POSIX(ON) ALL31(ON)/-p TCPCS -d
201.2.10.193'
```

  The PARM field can be `-c` for create, `-a` for create with affinity, or `-d` for delete. This example creates DVIPA 201.2.10.193 for the TCP/IP named TCPCS. After intermediate program is completed (and the comment character is removed), the DVIPA is deleted.

- On system 1, deactivate DVIPA 201.2.10.23 by issuing the following command on the MVS console:

```
V TCPIP,TCPCS,SYSPLEX,DEACTIVATE,DVIPA=201.2.10.23
```

Verifying the DVIPAs in a sysplex

The following display command can be used to display dynamic VIPAs in a sysplex:

```
d tcpip,tcpname,sysplex,vipadyn
```

In the following example taken from stack TCPCS, the ORIGIN lines show that 201.2.10.11, 201.2.10.12, 201.2.10.14, 201.2.10.15, DVIPA1, and DVIPA2 were all created by VIPADEFINE on this stack. 201.2.10.13, 201.2.10.21, and 201.2.10.22 were created by VIPABACKUP statements. (Note that the deactivated DVIPA 201.2.10.23 does not appear in this display.)

The ORIGIN line indicates how the DVIPA is configured on the stack specified by `tcpname`. Each stack (TCPNAME) for each system (MVSNAME) is shown with its status (STATUS). Two other status values not shown in the following example are:

QUIESCING
: This DVIPA has been deactivated, or it was a target for distribution and has been removed as a target. However, it is still servicing one or more connections for this DVIPA. The DVIPA will be removed when all connections complete.

MOVING
: This DVIPA was active on this stack and has been moved to another stack. Connections on this stack for this DVIPA prior to the move will still be serviced by this stack until completion.

The rank (RANK) indicates which of the stacks is eligible to take over if the stack on which the DVIPA is active stops. The stack with the highest rank is the one that will take over the DVIPA.

```
d tcpip,tcpcs,sysplex,vipadyn
EZZ8260I SYSPLEX CS V1R8 711
VIPA DYNAMIC DISPLAY FROM TCPCS    AT MVS004
LINKNAME: VIPLC9020A0B
IPADDR/PREFIXLEN: 201.2.10.11/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS2   MVS004   BACKUP 100  DEST
  TCPCS3   MVS005   BACKUP 010  DEST
  TCPCS6   MVS005   ACTIVE      DEST
LINKNAME: VIPLC9020A0C
IPADDR/PREFIXLEN: 201.2.10.12/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   BACKUP 075  DEST
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.13
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
  TCPCS    MVS004   BACKUP 100
  TCPCS3   MVS005   BACKUP 010
LINKNAME: VIPLC9020A0E
IPADDR/PREFIXLEN: 201.2.10.14/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   ACTIVE      DEST
LINKNAME: VIPLC9020A0F
IPADDR/PREFIXLEN: 201.2.10.15/28
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.21
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
```

```
  TCPCS3   MVS005   ACTIVE
  TCPCS2   MVS004   BACKUP 100
  TCPCS    MVS004   BACKUP 080
IPADDR: 201.2.10.22
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS    MVS004   BACKUP 080
  TCPCS2   MVS004   BACKUP 075
INTFNAME: DVIPA1
IPADDR: 2001:0DB8:1::1
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA2
IPADDR: 2001:0DB8:2::2
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA3
IPADDR: 2001:0DB8:3::3
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
INTFNAME: DVIPA4
IPADDR: 2001:0DB8:4::4
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
32 OF 32 RECORDS DISPLAYED
```

TCPCS2, TCPCS3, and TCPCS6 all display the same information about all the DVIPAs. ORIGIN fields are displayed for the DVIPAs that are configured on this stack.

```
d tcpip,tcpcs2,sys,vipad
EZZ8260I SYSPLEX CS V1R8 714
VIPA DYNAMIC DISPLAY FROM TCPCS2   AT MVS004
IPADDR: 201.2.10.11
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS2   MVS004   BACKUP 100  DEST
  TCPCS3   MVS005   BACKUP 010  DEST
  TCPCS6   MVS005   ACTIVE      DEST
IPADDR: 201.2.10.12
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   BACKUP 075  DEST
  TCPCS3   MVS005   BACKUP 010
LINKNAME: VIPLC9020A0D
IPADDR/PREFIXLEN: 201.2.10.13/26
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
  TCPCS    MVS004   BACKUP 100
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.14
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.15
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
```

```
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.21
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS2   MVS004   BACKUP 100
  TCPCS    MVS004   BACKUP 080
IPADDR: 201.2.10.22
  ORIGIN: VIPABACKUP
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS    MVS004   BACKUP 080
  TCPCS2   MVS004   BACKUP 075
INTFNAME: DVIPA1
IPADDR: 2001:0DB8:1::1
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA2
IPADDR: 2001:0DB8:2::2
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA3
IPADDR: 2001:0DB8:3::3
  ORIGIN: VIPADEFINE
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
INTFNAME: DVIPA4
IPADDR: 2001:0DB8:4::4
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
32 OF 32 RECORDS DISPLAYED
```

In the following example, TCPCS6 knows about the DVIPAs on the other stacks. There are no DVIPAs configured on TCPCS6, thus, no ORIGIN fields displayed.

```
d tcpip,tcpcs6,sys,vipad
EZZ8260I SYSPLEX CS V1R8 904
VIPA DYNAMIC DISPLAY FROM TCPCS6   AT MVS005
IPADDR: 201.2.10.11
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS2   MVS004   BACKUP 100  DEST
  TCPCS3   MVS005   BACKUP 010  DEST
  TCPCS6   MVS005   ACTIVE      DEST
IPADDR: 201.2.10.12
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   BACKUP 075  DEST
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.13
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
  TCPCS    MVS004   BACKUP 100
  TCPCS3   MVS005   BACKUP 010
IPADDR: 201.2.10.14
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      DIST
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.15
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
```

```
  TCPCS2   MVS004   ACTIVE      DEST
IPADDR: 201.2.10.21
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS2   MVS004   BACKUP 100
  TCPCS    MVS004   BACKUP 080
IPADDR: 201.2.10.22
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
  TCPCS    MVS004   BACKUP 080
  TCPCS2   MVS004   BACKUP 075
INTFNAME: DVIPA1
IPADDR: 2001:0DB8:1::1
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA2
IPADDR: 2001:0DB8:2::2
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS    MVS004   ACTIVE      BOTH
  TCPCS6   MVS005   ACTIVE      DEST
  TCPCS3   MVS005   ACTIVE      DEST
  TCPCS2   MVS004   ACTIVE      DEST
INTFNAME: DVIPA3
IPADDR: 2001:0DB8:3::3
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS2   MVS004   ACTIVE
INTFNAME: DVIPA4
IPADDR: 2001:0DB8:4::4
  TCPNAME  MVSNAME  STATUS RANK DIST
  -------- -------- ------ ---- ----
  TCPCS3   MVS005   ACTIVE
32 OF 32 RECORDS DISPLAYED
```

Using Netstat support to verify dynamic VIPA configuration

You can use the Netstat VIPADCFG/-F report option to display the dynamic VIPA configuration for a particular TCP/IP stack. Use the DETAIL modifier on the distributing stack to verify all the options configured on the VIPADISTRIBUTE statements.

Guideline: Use the Netstat CONFIG/-f report option to verify the rest of the stack configuration, including SOURCEVIPA, TCPSTACKSOURCEVIPA, and SYSPLEXROUTING; use the Netstat SRCIP/-J report option to verify the SRCIP profile statement for particular jobs or destinations.

The dynamic VIPA information section is displayed only when there are DVIPAs configured on this stack. The VIPA Range section, displayed only if a VIPARANGE statement was processed in this stack's initial profile (or in a data set referenced by a VARY TCPIP,,OBEYFILE command), indicates only that a range was configured. It does not indicate whether any ioctl or BIND has actually created a DVIPA in the specified range. The VIPA distribute section is displayed only if there are VIPADISTRIBUTE statements configured on this stack. The deactivated dynamic VIPA information section is displayed only when there are DVIPAs that have been deactivated on this stack.

Netstat CONFIG/-F DETAIL output from stack TCPCS:

```
# netstat -p tcpcs -F DETAIL
MVS TCP/IP NETSTAT CS V1R12       TCPIP Name: TCPCS           18:23:46
Dynamic VIPA Information:

  VIPA Backup:
    IpAddr/PrefixLen: 201.2.10.13
      Rank: 000100  Moveable:
    IpAddr/PrefixLen: 201.2.10.21
      Rank: 000080  Moveable:
    IpAddr/PrefixLen: 201.2.10.22
      Rank: 000080  Moveable:

  VIPA Define:
```

```
    IpAddr/PrefixLen: 201.2.10.11/28
      Moveable: Immediate
    IpAddr/PrefixLen: 201.2.10.12/28
      Moveable: Immediate
    IpAddr/PrefixLen: 201.2.10.14/28
      Moveable: Immediate
    IpAddr/PrefixLen: 201.2.10.15/28
      Moveable: Immediate
    IntfName: DVIPA1
      IpAddr: 2001:0DB8:1::1
        Moveable: Immediate
    IntfName: DVIPA2
      IpAddr: 2001:0DB8:2::2
        Moveable: Immediate

  VIPA Range:
    IpAddr/PrefixLen: 201.2.10.192/26
      Moveable: NonDisr
    IntfName: VRANGE1
      IpAddr/PrefixLen: 2001:0DB8:3::1/100
        Moveable: NonDisr

  VIPA Distribute:
    Dest:        201.2.10.11..20
      DestXCF:   ALL
        SysPt:   Yes  TimAff: No    Flg: ServerWLM
        OptLoc:  No
        ProcXcost:
          zAAP: 001  zIIP: 001
        ILWeighting: 0
    Dest:        201.2.10.11..21
      DestXCF:   ALL
        SysPt:   Yes  TimAff: No    Flg: ServerWLM
        OptLoc:  No
        ProcXcost:
          zAAP: 001  zIIP: 001
        ILWeighting: 2
    Dest:        201.2.10.12..20
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: BaseWLM
        OptLoc:  No
        ProcType:
          CP: 01  zAAP: 00  zIIP: 00
    Dest:        201.2.10.12..21
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: BaseWLM
        OptLoc:  No
        ProcType:
          CP: 01  zAAP: 00  zIIP: 00
    Dest:        201.2.10.14..n/a
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: Roundrobin
    Dest:        201.2.10.15..5000
      DestXCF:   193.9.200.2
        SysPt:   No   TimAff: No    Flg: WeightedActive
        OptLoc:  No   Weight: 10
    Dest:        201.2.10.15..23
      DestXCF:   ALL
        SysPt:   No   TimAff: 30    Flg: BaseWLM
        OptLoc:  No
        ProcType:
          CP: 20  zAAP: 80  zIIP: 00
    DestIntf:    DVIPA1
      Dest:      2001:0DB8:1::1..n/a
        DestXCF: ALL
        SysPt:   Yes  TimAff: No    Flg: BaseWLM
        OptLoc:  No
        ProcType:
          CP: 01  zAAP: 00  zIIP: 00
    DestIntf:    DVIPA2
      Dest:      2001:0DB8:2::2..23
        DestXCF: ALL
        SysPt:   No   TimAff: 45    Flg: Roundrobin

Deactivated Dynamic VIPA Information:

  VIPA Define:
    IpAddr/PrefixLen: 201.2.10.23/28
      Moveable: Immediate

  VIPA Distribute:
    Dest:        201.2.10.23..4000
```

```
      DestXCF:   ALL
        SysPt:   No   TimAff: No    Flg: BaseWLM
        OptLoc:  No
        ProcType:
          CP: 01  zAAP: 00  zIIP: 00
```

On stack TCPCS2 from the console:

```
d tcpip,tcpcs2,net,vipadcfg
EZD0101I NETSTAT CS V1R8 TCPCS2 721
DYNAMIC VIPA INFORMATION:
  VIPA BACKUP:
    IPADDR/PREFIXLEN: 201.2.10.11
      RANK: 000100  MOVEABLE:
    IPADDR/PREFIXLEN: 201.2.10.12
      RANK: 000075  MOVEABLE:
    IPADDR/PREFIXLEN: 201.2.10.21
      RANK: 000100  MOVEABLE:
    IPADDR/PREFIXLEN: 201.2.10.22
      RANK: 000075  MOVEABLE:
  VIPA DEFINE:
    IPADDR/PREFIXLEN: 201.2.10.13/26
      MOVEABLE: IMMEDIATE
    INTFNAME: DVIPA3
      IPADDR: 2001:0DB8:3::3
        MOVEABLE: IMMEDIATE
  VIPA RANGE:
    IPADDR/PREFIXLEN: 201.2.10.192/26
      MOVEABLE: NONDISR
END OF THE REPORT
```

On stack TCPCS3 from the console:

```
d tcpip,tcpcs3,net,vipadcfg
EZD0101I NETSTAT CS V1R8 TCPCS3 907
DYNAMIC VIPA INFORMATION:
  VIPA BACKUP:
    IPADDR/PREFIXLEN: 201.2.10.11
      RANK: 000010  MOVEABLE:
    IPADDR/PREFIXLEN: 201.2.10.12
      RANK: 000010  MOVEABLE:
    IPADDR/PREFIXLEN: 201.2.10.13
      RANK: 000010  MOVEABLE:
  VIPA DEFINE:
    IPADDR/PREFIXLEN: 201.2.10.21/26
      MOVEABLE: IMMEDIATE
    IPADDR/PREFIXLEN: 201.2.10.22/26
      MOVEABLE: IMMEDIATE
    INTFNAME: DVIPA4
      IPADDR: 2001:0DB8:4::4
        MOVEABLE: IMMEDIATE
  VIPA RANGE:
    IPADDR/PREFIXLEN: 201.2.10.192/26
      MOVEABLE: NONDISR
END OF THE REPORT
```

On stack TCPCS6 from the console:

```
d tcpip,tcpcs6,net,vipadcfg
EZD0101I NETSTAT CS V1R8 TCPCS6 910
END OF THE REPORT
```

The Netstat VIPADyn/-v report option displays all the dynamic VIPAs available to this stack, as shown in the following examples. (Note that deactivated DVIPAs do not appear in this report.)

Netstat VIPADyn/-v output from stack TCPCS:

```
# netstat -p tcpcs -v
MVS TCP/IP NETSTAT CS V1R8       TCPIP Name: TCPCS           18:32:26
     IpAddr/PrefixLen: 201.2.10.11/28
       Status:  Active     Origin: VIPADefine       DistStat: Dist/Dest
       ActTime: 03/02/2005 16:45:20
     IpAddr/PrefixLen: 201.2.10.12/28
       Status:  Active     Origin: VIPADefine       DistStat: Dist
       ActTime: 03/02/2005 16:45:20
     IpAddr/PrefixLen: 201.2.10.13/26
```

```
       Status:  Backup     Origin: VIPABackup       DistStat:
       ActTime: n/a
     IpAddr/PrefixLen: 201.2.10.14/28
       Status:  Active     Origin: VIPADefine       DistStat: Dist
       ActTime: 03/02/2005 16:45:20
     IpAddr/PrefixLen: 201.2.10.15/28
       Status:  Active     Origin: VIPADefine       DistStat: Dist/Dest
       ActTime: 03/02/2005 16:45:20
     IpAddr/PrefixLen: 201.2.10.21/26
       Status:  Backup     Origin: VIPABackup       DistStat:
       ActTime: n/a
     IpAddr/PrefixLen: 201.2.10.22/26
       Status:  Backup     Origin: VIPABackup       DistStat:
       ActTime: n/a
     IntfName: DVIPA1
       IpAddr: 2001:0DB8:1::1
         Status:  Active     Origin: VIPADefine       DistStat: Dist/Dest
         ActTime: 03/02/2005 16:45:20
     IntfName: DVIPA2
       IpAddr: 2001:0DB8:2::2
         Status:  Active     Origin: VIPADefine       DistStat: Dist/Dest
         ActTime: 03/02/2005 16:45:20
```

On stack TCPCS2 from the console:

```
d tcpip,tcpcs2,net,vipadyn
EZD0101I NETSTAT CS V1R8 TCPCS2 731
     IPADDR/PREFIXLEN: 201.2.10.11/28
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.12/28
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.13/26
       STATUS:  ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.14/28
       STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.15/28
       STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.21/26
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
       ActTime: n/a
     IPADDR/PREFIXLEN: 201.2.10.22/26
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
       ActTime: n/a
     INTFNAME: DVIPA1
       IPADDR: 2001:0DB8:1::1
         STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
         ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA2
       IPADDR: 2001:0DB8:2::2
         STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
         ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA3
       IPADDR: 2001:0DB8:3::3
         STATUS:  ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
         ActTime: 03/02/2005 16:45:20
     10 OF 10 RECORDS DISPLAYED
```

On stack TCPCS3 from the console:

```
d tcpip,tcpcs3,net,vipadyn
EZD0101I NETSTAT CS V1R8 TCPCS3 913
     IPADDR/PREFIXLEN: 201.2.10.11/28
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.12/28
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
       ActTime: n/a
     IPADDR/PREFIXLEN: 201.2.10.13/26
       STATUS:  BACKUP     ORIGIN: VIPABACKUP       DISTSTAT:
       ActTime: n/a
     IPADDR/PREFIXLEN: 201.2.10.15/28
       STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.21/26
```

```
       STATUS:  ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.22/26
       STATUS:  ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
       ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA1
       IPADDR: 2001:0DB8:1::1
         STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
         ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA2
       IPADDR: 2001:0DB8:2::2
         STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
         ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA4
       IPADDR: 2001:0DB8:4::4
         STATUS:  ACTIVE     ORIGIN: VIPADEFINE       DISTSTAT:
         ActTime: 03/02/2005 16:45:20
     9 OF 9 RECORDS DISPLAYED
```

On stack TCPCS6 from the console:

```
d tcpip,tcpcs6,net,vipadyn
EZD0101I NETSTAT CS V1R8 TCPCS6 916
     IPADDR/PREFIXLEN: 201.2.10.11/28
       STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     IPADDR/PREFIXLEN: 201.2.10.15/28
       STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
       ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA1
     IPADDR: 2001:0DB8:1::1
         STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
         ActTime: 03/02/2005 16:45:20
     INTFNAME: DVIPA2
       IPADDR: 2001:0DB8:2::2
         STATUS:  ACTIVE     ORIGIN:                  DISTSTAT: DEST
         ActTime: 03/02/2005 16:45:20
     4 OF 4 RECORDS DISPLAYED
```

Verifying sysplex distributor workload

You can use the Netstat VDPT/-O and VCRT/-V report options to verify sysplex distributor workload. See z/OS Communications Server: IP System Administrator's Commands for more information on these commands.

Use the Netstat VDPT/-O report option with the DETAIL modifier on the distributing stack to confirm that there are target stacks available with server applications ready. This display will show only target stacks that are currently up and have joined the sysplex. The RDY field indicates how many, if any, applications the target TCP/IP, identified by its DestXCF Addr, has bound to DPort. If none, then this target TCP/IP will not receive any connection workload. The TotalConn field indicates how many connections this distributing TCP/IP has forwarded to the target TCP/IP. The ActiveConn field indicates how many connections the distributing TCP/IP has forwarded to the target TCP/IP that are currently active.

Tip: TotalConn is an historical count and will wrap.

The following Netstat display command on the distributing stack, TCPCS, shows which target stacks are available with the server applications ready.

```
NETSTAT VDPT DETAIL

MVS TCP/IP NETSTAT CS V1R8       TCPIP Name: TCPCS           15:37:51
Dynamic VIPA Destination Port Table:
Dest:        201.2.10.11..21
  DestXCF:   201.1.10.15
  TotalConn: 0000000000  Rdy: 001  WLM: 01 TSR: 075
  Flg: Dynamic, Roundrobin, Inactive
    TCSR: 100 CER: 075 SEF: 100
    Abnorm: 00         Health: 100
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 01
Dest:        201.2.10.12..4011
  DestXCF:   201.1.10.15
```

```
  TotalConn: 0000000000  Rdy: 001  WLM: 02 TSR: 100
  Flg: Roundrobin
    TCSR: 100 CER: 100 SEF: 100
    Abnorm: 00         Health: 100
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 00
    QosPlcAct: paPRD-SD-7-INTR-DONTCARE
      W/Q: 02
Dest:        201.2.10.12..4011
  DestXCF:   201.2.10.10
  TotalConn: 0000000000  Rdy: 001  WLM: 02 TSR: 050
  Flg: Roundrobin, Inactive
    TCSR: 067 CER: 075 SEF: 100
    Abnorm: 00         Health: 100
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 01
    QosPlcAct: paPRD-SD-7-INTR-DONTCARE
      W/Q: 02
Dest:        201.2.10.13..243
  DestXCF:   201.3.10.16
  TotalConn: 0000000000  Rdy: 001  WLM: 02 TSR: 085
  Flg: BaseWLM
    TCSR: 100 CER: 095 SEF: 090
    Abnorm: 00         Health: 100
    Weight 50
      Raw           CP:  50 zAAP: 00 zIIP: 63
      Proportional  CP:  50 zAAP: 00 zIIP: 00
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 02
Dest:        201.2.10.23..4000
  DestXCF:   201.3.10.16
  TotalConn: 0000000000  Rdy: 001  WLM: 15 TSR: 100
  Flg: BaseWLM
    TCSR: 100 CER: 100 SEF: 100
    Abnorm: 00         Health: 100
    Weight 60
      Raw           CP:  50 zAAP: 00 zIIP: 63
      Proportional  CP:  10 zAAP: 00 zIIP: 50
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 02
DestIntf:
  Dest:        2001:0db8::522:f103..20
    DestXCF:   2001:0db8::943:f003
    TotalConn: 0000000000  Rdy: 001  WLM: 01 TSR: 094
    Flg: BaseWLM Local
    TCSR: 099 CER: 098 SEF: 097
    Abnorm: 00         Health: 100
    Weight 4
      Raw           CP:  20 zAAP: 00 zIIP: 00
      Proportional  CP:  04 zAAP: 00 zIIP: 00
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 00
DestIntf:
  Dest:        2001:0db8::522:f103..21
    DestXCF:   2001:0db8::943:f003
    TotalConn: 0000000000  Rdy: 001  WLM: 01 TSR: 100
    Flg: Roundrobin
    TCSR: 100 CER: 100 SEF: 100
    Abnorm: 00         Health: 100
  ActiveConn: 0000000000
    QosPlcAct: *DEFAULT*
      W/Q: 01
```

- The target stack is identified by its dynamic XCF address (DESTXCF).
- The RDY field indicates how many applications on that target stack have bound to the DEST port.
- TOTALCONN is the number of all connections the distributing stack, TCPCS, has routed to the target stack.
- The ACTIVECONN value is the number of currently active connections that the distributing stack, TCPCS, has routed to the target stack. The active connections include all TCP connections that the target TCP/IP stack is still aware of, even if the connections are in the process of terminating. For example, connections in FIN_WAIT2 or TIMEWAIT states are also included in this count.

- The ABNORM value is the rate of abnormal terminations for this server. This field affects the WLM weight and the operation of the OPTLOCAL distributor setting.
- The HEALTH value is the health indicator for this server. This field affects the WLM weight and the operation of the OPTLOCAL distributor setting.
- WLM is the normalized workload manager weight value for the target TCP/IP stack.
- Weight is the raw composite weight for the target TCP/IP stack. The composite weight is based on the application's general CPU usage, the System z Application Assist Processor (zAAP) usage, and the System z Integrated Information Processor (zIIP) usage, as shown in Table 24 on page 429.

Table 24. Weight determination

Processor	DISTMETHOD BASEWLM	DISTMETHOD SERVERWLM
CP	The Raw value is the raw WLM system general CP weight. The Proportional value is the Raw value modified by the expected CP usage proportion configured on the VIPADISTRIBUTE PROCTYPE statement for this application.	The Raw value is the raw WLM server-specific general CP weight. The Proportional value is the Raw value modified by the proportion of CP capacity that is currently being consumed by the application's workload, as compared to the other processors (zIIP and zAAP).
zAAP	The Raw value is the raw WLM system zAAP weight. The Proportional value is the Raw value modified by the expected zAAP usage proportion configured on the VIPADISTRIBUTE PROCTYPE statement for this application.	The Raw value is the raw WLM server-specific zAAP weight. The Proportional value is the Raw value modified by the proportion of zAAP capacity that is currently being consumed by the application's workload, as compared to the other processors (CP and zIIP).
zIIP	The Raw value is the raw WLM system zIIP weight. The Proportional value is the Raw value modified by the expected zIIP usage proportion configured on the VIPADISTRIBUTE PROCTYPE statement for this application.	The Raw value is the raw WLM server-specific zIIP weight. The Proportional value is the Raw value modified by the proportion of zIIP capacity that is currently being consumed by the application's workload, as compared to the other processors (CP and zAAP).

- FLG is a flag field indicating how the connection was distributed. For example, the value DYNAMIC indicates this DVIPA is being distributed by the dynamic port assignment function.
- The target server responsiveness (TSR) field indicates how well a target server is accepting TCP connection setup requests. A value of 100 indicates that the target server is accepting all TCP connection setup requests successfully. A value of 0 indicates that the target server is accepting no TCP connection setup requests successfully. A value between 100 and 0 indicates the relative success rate of TCP connection setup requests for this target server.

When weighted active distribution is being used, the configured active connection weight for each target is shown as the WLM value in the Netstat VDPT/-O display. By using the DETAIL modifier on the display, the active connection counts (ActiveConn) are shown. The distributor balances incoming connection requests across the targets with a goal of having the number of active connections on each target proportionally equivalent to the configured active connection weight of each target.

```
d tcpip,tcpcs,net,vdpt,detail
DEST:        201.2.10.15..5000
  DESTXCF:   193.9.200.2
  TOTALCONN: 0000021015  RDY: 001  WLM: 10   TSR: 100
  FLG: WeightedActive
    TCSR: 100 CER: 100 SEF: 100
    Abnorm: 00         Health: 100
```

```
    ActiveConn: 0000001555

    QosPlcAct: *DEFAULT*
      W/Q: 00

DEST:        201.2.10.15..5000
  DESTXCF:   193.9.200.2
  TOTALCONN: 0000044015  RDY: 002  WLM: 20   TSR: 100
  FLG: WeightedActive
    TCSR: 100 CER: 100 SEF: 100
    Abnorm: 00         Health: 100

    ActiveConn: 0000003100

    QosPlcAct: *DEFAULT*
      W/Q: 00
```

The following Netstat display command on the distributing stack displays all current connections being distributed by TCPCS.

```
d tcpip,tcpcs,net,vcrt
EZD0101I NETSTAT CS V1R8 TCPCS 844
DYNAMIC VIPA CONNECTION ROUTING TABLE:
DEST:      201.2.10.11..21
  SOURCE:  193.9.200.5..1029
  DESTXCF: 193.9.200.1
DEST:      201.2.10.11..21
  SOURCE:  193.9.200.8..1050
  DESTXCF: 193.9.200.2
DEST:      201.2.10.11..21
  SOURCE:  193.9.200.11..1079
  DESTXCF: 193.9.200.3
DEST:      201.2.10.12..21
  SOURCE:  193.9.200.9..1030
  DESTXCF: 193.9.200.2
DEST:      2001:0DB8:1::1..21
  SOURCE:  2001:0DB8::151:0006..1038
  DESTXCF: 2001:0DB8::151:0003
END OF THE REPORT
```

For more details, see z/OS Communications Server: IP System Administrator's Commands.

Dynamic VIPAs and routing protocols

With dynamic VIPAs, IP addresses may move from one stack to another. These changes need to be communicated to the network. Therefore, dynamic routing should be implemented when dynamic VIPAs are being used.

IPv4 considerations for OMPROUTE

The names of dynamic VIPA interfaces are assigned dynamically by the stack when a dynamic VIPA interface is created. Therefore, the Name field that is set on the Interface or OSPF_Interface statement for a dynamic VIPA is ignored by OMPROUTE.

Define an Interface or OSPF_Interface definition for every dynamic VIPA address that a host might own. Because there could be many interfaces, more wildcard capabilities are added to OMPROUTE, for dynamic VIPA interfaces only.

Ranges of dynamic VIPA interfaces can be defined by using the subnet mask parameter on the OSPF_Interface or Interface statement. The range that is defined is all the IP addresses that fall within the subnet that is defined by the mask and the IP address. The IP address parameter must be the subnet number of the range that is being defined, not a host address within that range. The following example defines a range of dynamic VIPA addresses from 10.138.165.81 to 10.138.165.94:

```
OSPF_Interface
   IP_address = 10.138.165.80
   Name = dummy_name (see tip)
   Subnet_mask = 255.255.255.240;
```

Tips:

- The Name parameter is required and must be unique, but it is not used for dynamic VIPAs.
- When you are defining ranges, it is not necessary or desirable to code a destination address. OMPROUTE automatically sets the destination address of a dynamic VIPA to its IP address.
- There is nothing in the interface definition statements that informs OMPROUTE that a particular interface definition statement is for a dynamic VIPA or a range of dynamic VIPAs. Rather, OMPROUTE learns this information from the stack when these interfaces are created or taken over.
- Because dynamic VIPAs can move between z/OS hosts, configure a router ID that specifies a physical interface or a static VIPA and not a dynamic VIPA.

OMPROUTE cannot build an advertisement with a size that exceeds the largest IP packet size, which is 64K. When this limit is reached, the following message is displayed:

```
EZZ7967I ADVERTISEMENT DISCARDED, OVERFLOWS BUFFER: LS
   TYPE x ID x.x.x.x ORG y.y.y.y
```

Result: If this limit is reached with a router link-state advertisement (LSA), OMPROUTE cannot send the router LSA and other hosts cannot calculate routes to destinations (for example, VIPAs) owned by the stack on which OMPROUTE is running. OMPROUTE ends if it encounters this condition. If OMPROUTE cannot send its router LSA, it is useless as a router.

If an LSA exceeds the maximum transmission unit (MTU) size on its network, the LSA is fragmented. This fragmentation is not a problem for OMPROUTE, but some routers cannot accept fragmented LSAs.

Tip: If you are experiencing adjacency problems with routers and have many interfaces on your stack, check for fragmentation of the OMPROUTE router LSA.

Normally, large router LSAs are not a problem, as the size of LSAs seldom exceeds 64K, or even network MTU sizes. However, if many VIPA or dynamic VIPA interfaces are defined on a host, router LSA size can become a consideration. The size of the router LSA includes 52 bytes for headers, plus the number of bytes required to advertise the interfaces that the host owns. The number of bytes required for each interface is:

VIPA
12 bytes, plus 12 bytes for each VIPA subnet

Tip: If many VIPAs are in different subnets on the host, the size of the router LSA can be reduced by suppressing VIPA subnet routes by coding the Advertise_VIPA_Routes parameter on the OSPF_INTERFACE statement for the VIPA interfaces.

Point-to-point
24 bytes

Point-to-multipoint
12 bytes, plus 12 bytes for each neighbor on the interface

All other types
12 bytes

IPv4 considerations for Routing Information Protocol

If using Routing Information Protocol (RIP) services and Host Route advertising is not supported by adjacent routers (that is, inability to learn host routes), the following restrictions for VIPA addresses must be applied to benefit from fault tolerance support:

- If you use subnetting and VIPA addresses are in the same network as the physical IP addresses, the subnetwork portion of any VIPA addresses must not be the subnetwork portion of any physical IP addresses in the network. In this case, assign a new subnetwork for the VIPA address.
- If subnetting is not used on any physical interface, the network portion of any VIPA addresses must not be the network portion of any physical IP addresses in the network. In this case, assign a new network for the VIPA address, preferably a class C network address.

If using RIP services and Host Route advertising is supported by adjacent routers, the network or subnetwork portions of VIPA addresses can be the same across multiple z/OS TCP/IP stacks in the network. To enable Host Route advertising in OMPROUTE, configure RIP_Interface Send_Host_Routes=YES.

IPv6 considerations

For IPv6, all interfaces are defined to OMPROUTE by name only, and name wildcards are supported. This greatly simplifies definition considerations for OMPROUTE and VIPA interfaces.

To define IPv6 interfaces to OMPROUTE, you must know their names. Unlike IPv4, in which VIPA link names are generated by the stack, you specify link names to TCP/IP when defining IPv6 dynamic VIPA interfaces. These link names should be used when defining the same interfaces to OMPROUTE.

Tip: Using consistent naming conventions and name wildcards can make this task easier. For example, if you require that all IPv6 VIPA link names start with the letter V, and do not allow other interface types to have link names starting with V, a single definition can define all IPv6 VIPA interfaces to OMPROUTE as follows:

```
IPV6_OSPF_INTERFACE
      NAME=V*;
```

Unlike IPv4, there are no size limitations on the number of IPv6 dynamic VIPA addresses that OMPROUTE can advertise. In IPv6 OSPF, OMPROUTE will advertise all host and prefix addresses associated with owned interfaces, including VIPAs and dynamic VIPAs. In IPv6 RIP, sending and receiving of host routes should be turned on to allow host addresses associated with VIPAs to be advertised.

Chapter 8. TCP/IP in a sysplex

The increasing demands of network servers, and in particular System z servers, has led to the creation of different techniques to address performance requirements when a single server is not capable of providing the availability and scalability demands placed on it by its clients. Specifically, network solutions make use of what is referred to as the clustering technique, whereby multiple servers are associated together into a cluster to provide sufficient processing power and availability characteristics to handle the demands of the clients.

In the scope of this topic, this cluster functionality is provided by the sysplex. That is, the sysplex provides the necessary capability to cluster together a number of System z servers that cooperate with one another to deliver the processing power needed to service the demands required of a particular service environment.

Solutions that use the clustering approach to increase server availability and processing capability attempt to provide mechanisms by which they ensure the viability of the cluster in an environment containing a large number of clients generating a potentially high number of requests. To do so, the cluster technique can provide for two main objectives, high availability and load balancing. In some cases, clustering techniques address only high availability, as is the case with dynamic VIPA that provides for availability in spite of potential TCP/IP stack or z/OS image failures. In other cases, the intent is to provide for both high availability and load balancing, as is done by sysplex distributor.

In general, load balancing refers to the ability to use different systems within the cluster simultaneously, thereby taking advantage of the additional computational function of each. Further, clustering techniques addressing load balancing lead to other system requirements, such as that of a single systemwide image (one identity by which clients access the system), horizontal growth, and ease of management.

The traditional view of a single server has been primarily a single machine with perhaps a few network interfaces (IP addresses). This tends to lead to many potential points of failure within the server: the machine itself (hardware), the operating system (including TCP/IP stack) kernel executing on the machine, or a network interface (and the IP address associated with it). Static Virtual IP Addresses (VIPAs) exclude the network interface as a point of failure while dynamic VIPAs additionally aid with server (image) or kernel failure. In this way, high availability is seen as the availability of the entire server cluster and the service it provides. Furthermore, VIPAs can be used in conjunction with the sysplex distributor load balancing solution.

Clustering techniques that address the load balancing of connections requests also typically provide for some high availability. That is, these techniques dispatch connections to target servers and can exclude failed servers from the list of target servers that can receive connections. In this way, the dispatching function avoids routing connections and requests to a server incapable of satisfying such requests.

Load balancing is the ability for a cluster to distribute workload evenly (or based on some policy) to target servers comprising the cluster. Usually, this load balancing is measured by some notion of perceived load on each of the target servers. This topic describes load balancing using the sysplex distributor, which identifies the target System z servers willing to receive client connections based on some specification.

By providing load balancing, clustering techniques must also provide for other system requirements in addition to the dispatching of connections. These include the ability to advertise some single systemwide image or identity so that clients can uniquely and easily identify the service. Additionally, clustering techniques should also provide for horizontal growth of the system and ease of management.

It is also sometimes useful to arrange the members in the sysplex into subsets of members that communicate through XCF only amongst themselves. This is called subplexing. Each member of a subplex joins a unique set of sysplex groups and communicates through XCF with only other members that have joined the same unique set of sysplex groups. Each TCP/IP stack can participate in one and only one subplex. For more information, see “Sysplex subplexing” on page 436.

Note: This information applies to both IPv4 and IPv6, unless otherwise noted.

© Copyright IBM Corp. 2000, 2022

Connectivity in a sysplex

With dynamic VIPAs, IP addresses may move from one stack to another. These changes need to be communicated to the network. Therefore, dynamic routing should be implemented when dynamic VIPAs are being used. See “Dynamic VIPAs and routing protocols” on page 431 for more detailed information.

Sysplex subplexing

A subplex is a subset of a sysplex that consists of all the members in the sysplex that communicate through cross-system coupling facility (XCF) groups with each other, and not with members outside the subset.

Defining multiple subplex scopes within a sysplex can be useful in scenarios where multiple networks with different security or functional attributes are attached to specific systems in the sysplex. For example, consider the scenario where some systems in the sysplex are connected to an internal enterprise network and other systems in the sysplex are connected to external networks that have different security attributes. In this example, you might want to be able to partition the sysplex into two subplexes from a sysplex network function perspective; one that includes the systems connected to the internal network and another for the systems that are connected to the external networks. By defining these separate subplexes, users can still exploit some of the sysplex network functions while preserving the isolation of the networks (that is, without automatically enabling dynamic XCF connectivity across the entire sysplex).

TCP/IP and VTAM subplex concepts and example

To enable partitioning of the TCP/IP stacks and VTAM nodes in a sysplex into subplexes, the values specified by the following VTAM start option and TCP/IP profile GLOBALCONFIG parameter are used as suffixes for XCF group names and coupling facility structure names.

- VTAM

  Use the XCFGRPID start option. For a description of sysplex subplexing and how the XCFGRPID start option modifies VTAM group names and VTAM structure names, see z/OS Communications Server: SNA Network Implementation Guide.
- TCP/IP

  Use the XCFGRPID parameter on the GLOBALCONFIG statement in the TCP/IP profile.

For TCP/IP, both the VTAM group ID suffix and the TCP group ID suffix can be used to modify the group name that a TCP/IP stack uses when it joins the sysplex. The group name used is in the form EZBT*vvtt*, where *vv* is the 2-digit VTAM group ID suffix specified on the XCFGRPID start option, and *tt* is the TCP group ID suffix specified on the GLOBALCONFIG statement in the TCP/IP profile. By using these suffixes to modify the XCF group name, the sysplex is effectively partitioned into subsets referred to as subplexes.

Only those TCP/IP stacks joining the unique group name generated using the specified suffixes are aware of each other in the sysplex; they communicate with each other through dynamic XCF. If the VTAM node is stopped and restarted with a new value for its XCFGRPID start option, the TCP/IP stacks using that VTAM node must be stopped and restarted to access the new VTAM suffix.

Figure 55 on page 437 shows an example of TCP/IP and VTAM subplexes.

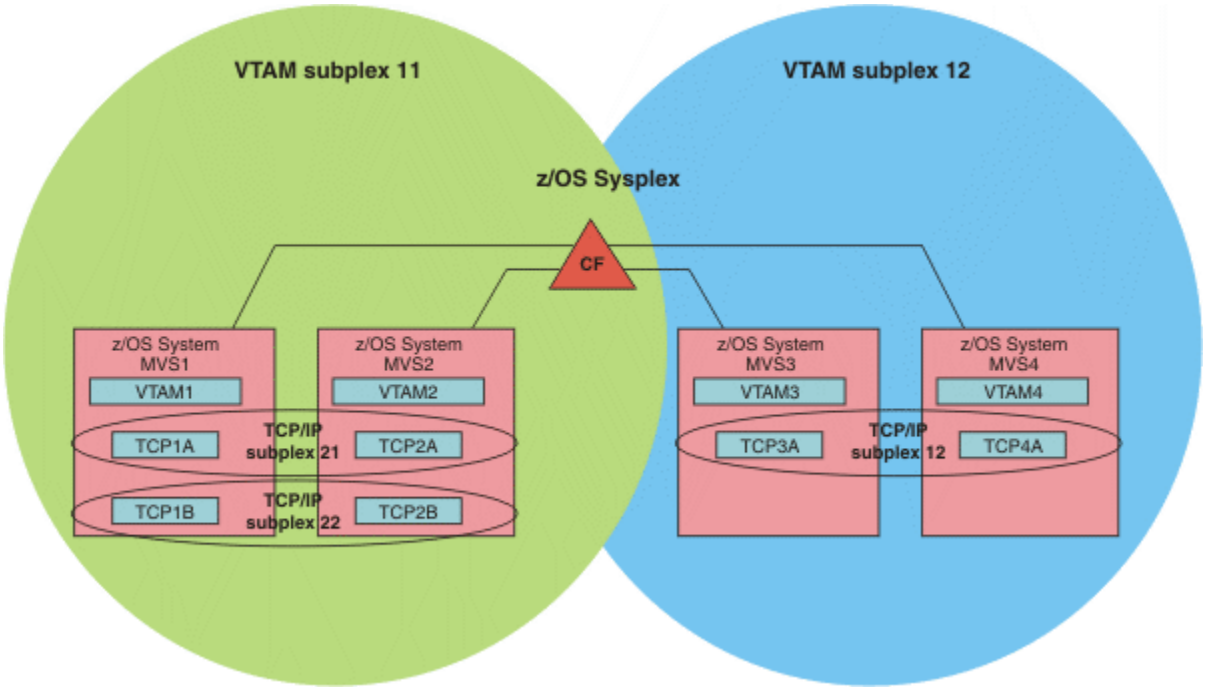


Figure 55. An example of TCP/IP and VTAM subplexes

In Figure 55 on page 437, the following subplexes are defined:

- Two VTAM subplexes (11 and 12) have been defined by starting VTAM1 and VTAM2 with a start option of XCFGRPID 11, and starting VTAM3 and VTAM4 with a start option of XCFGRPID 12.
- Three TCP/IP subplexes have been defined by starting TCP1A and TCP2A with the statement GLOBALCONFIG XCFGRPID 21, TCP1B and TCP2B with the statement GLOBALCONFIG XCFGRPID 22, and TCP3A and TCP4A with the statement GLOBALCONFIG XCFGRPID 12.

The following conditions exist in the sysplex shown in Figure 55 on page 437:

- TCP/IP stacks TCP1A and TCP2A join XCF group EZBT1121. TCP1A is aware of only TCP2A in the sysplex, and TCP2A is aware of only TCP1A in the sysplex.
- TCP/IP stacks TCP1B and TCP2B join XCF group EZBT1122. TCP1B is aware of only TCP2B in the sysplex, and TCP2B is aware of only TCP1B in the sysplex.
- TCP/IP stacks TCP3A and TCP4A join XCF group EZBT1212. TCP3A is aware of only TCP4A in the sysplex, and TCP4A is aware of only TCP3A in the sysplex.

Note: Even if TCP3A and TCP4A had been started with a GLOBALCONFIG 21 value, they would still be in a separate group from TCP1A and TCP2A, because they are in different VTAM subplexes.

Rule: Because TCP/IP stacks use VTAM XCF support for their dynamic XCF connectivity, TCP/IP subplexes cannot span multiple VTAM subplexes.

To preserve separation between the subplexes, the TCP and VTAM coupling facility structures accessed must also be unique for each subplex. For the TCP structures EZBDVIPA, used for SWSA support, and EZBEPORT, used for SYSPLEXPORTS support, this is accomplished by appending the VTAM and TCP XCF group ID suffixes to the end of the structure names, or EZBDVIPA*vvtt* and EZBEPORT*vvtt* respectively. Thus, in the example, TCP1A and TCP2A access EZBDVIPA1121 and EZBEPORT1121, while TCP1B and TCP2B access EZBDVIPA1122 and EZBEPORT1122.

To access the TCP structures, the structure names, including the suffixes, must be defined in the sysplex CFRM policy. In Figure 55 on page 437, the CFRM policy for the sysplex must have structure definitions for EZBDVIPA1121, EZBDVIPA1122, EZBDVIPA1212, EZBEPORT1121, EZBEPORT1122, and EZBEPORT1212, if all of the stacks expect to access these structures.

- If no VTAM suffix is specified, a value of CP is used for *vv* when creating the TCP sysplex group name, and a value of 01 is used for *vv* when creating the structure names.

- If a VTAM suffix is specified, but no TCP suffix is specified, a value of CS is used for *tt* when creating the TCP sysplex group name, and a null value is used for *tt* when creating the structure names. Thus, in Figure 55 on page 437, if TCP1B and TCP2B are started without an XCFGRPID parameter on their GLOBALCONFIG statements, the sysplex group they join is EZBT11CS, and the structures they can access are EZBDVIPA11 and EZBEPORT11.

To estimate the coupling facility size requirements for each of these structures, use the CFSizer website at http://www.ibm.com/systems/support/z/cfsizer/.

To isolate communication between different subplexes within a sysplex, dynamic XCF connectivity between TCP/IP stacks in different subplexes on separate MVS images is blocked. In addition, IUTSAMEH connectivity for dynamic XCF between stacks in different subplexes on the same MVS image is also blocked.

When using HiperSockets to establish dynamic XCF connectivity between TCP/IP stacks in the same CPC using the same HiperSockets CHPID, to preserve subplex isolation, specify the IQDVLANID parameter on the GLOBALCONFIG statement as follows:

- TCP/IP stacks in the same subplex that share the same CPC and specify the same HiperSockets CHPID must specify the same VLAN ID value on the IQDVLANID parameter.
- TCP/IP stacks in different subplexes that share the same CPC and specify the same HiperSockets CHPID must specify different VLAN ID values on the IQDVLANID parameter.
- If subplexing is not being used within the sysplex, do not specify the IQDVLANID parameter on the GLOBALCONFIG statement.

Static XCF, user-defined IUTSAMEH, or user-defined HiperSockets connectivity between TCP stacks in different TCP subplexes within the same VTAM subplex are not blocked, because you have control over the definition and activation of these connections.

No steps are explicitly taken to block any other network connectivity available through LANs and WANs that might be shared by different subplexes, including scenarios where the same OSA-Express adapter is shared across systems in different subplexes. If complete network isolation between subplexes is required, other mechanisms, such as physical network isolation or logical separation through firewall technologies, should be deployed.

Subplex support applies to TCP/IP and VTAM sysplex functions that make explicit use of XCF communications and coupling facility structures. For TCP/IP, this includes functions such as dynamic XCF, dynamic VIPAs, sysplex distributor, SYSPLEXPORTS, and SWSA. The scope of these functions is typically the entire sysplex. Automated domain name registration (ADNR) does not have explicit support for subplexing, but if ADNR is used in a subplexing environment, the scope of ADNR might be restricted to a subplex when the different subplexes are used to isolate network connectivity. In this case, an instance of ADNR must be started for each subplex in the sysplex that will use ADNR functions. ADNR needs to have affinity to one of the stacks in the subplex when ADNR is started on a system that is running multiple stacks. See adnrproc.sample for an example of the JCL.

In the remainder of this information, references to the sysplex or the sysplex group can refer to the entire sysplex and sysplex group EZBTCPCS, if subplexing is not being used, or to a specific subplex and a specific instance of the sysplex group EZBT*vvtt*, if subplexing is in effect.

Setting up a subplex

Setting up TCP/IP and VTAM subplexes within a sysplex requires some preparation.

Steps for preparing your sysplex for subplexing

You need to plan how many subplexes are needed, assign group IDs, ensure that all TCP/IP stacks and VTAM nodes in the sysplex are at z/OS V1R8 or later, and assign VLAN IDs as needed.

Before you begin

Steps 2 and 3 depend on the level of isolation you require for your subplexes within the sysplex. Those two steps are not needed when the following conditions are true:

- You do not use HiperSockets in support of dynamic XCF connectivity
- You do not need to isolate the subplexes for HiperSockets connectivity (if, for example, you need only to restrict sysplex distributor distribution between subplexes)
- You can configure your systems such that TCP/IP stacks that are in separate subplexes and are on the same CPC use a CHPID that is unique to the subplex

Requirement: A z890 GA2 or z990 GA2 hardware level is required to support VLAN IDs on HiperSockets.

Procedure

Perform the following steps to prepare your sysplex for subplexing.

1. Plan how many subplexes are needed and assign a TCP/IP XCF group ID and a VTAM XCF group ID for each anticipated subplex in the sysplex:
   - All TCP/IP stacks in the same TCP/IP subplex must have the same TCP/IP XCF group ID.
   - All VTAM nodes in the same VTAM subplex must have the same VTAM XCF group ID.
2. Ensure that all TCP/IP stacks and VTAM nodes in the sysplex are at z/OS V1R8 or later.
3. If HiperSockets is enabled and there are TCP/IP stacks in the sysplex that are on the same CPC and use the same HiperSockets CHPID, assign a VLAN ID to each anticipated subplex in the sysplex. Also, assign a VLAN ID to the default subplex (that is, the subset of sysplex TCP/IP stacks that uses the default sysplex group name, EZBTCPCS). All TCP/IP stacks in the same subplex must have the same VLAN ID, if they are on the same CPC and use the same HiperSockets CHPID.

Steps for partitioning a set of TCP/IP stacks in a sysplex into a subplex

You can partition a set of TCP/IP stacks in a sysplex into a subplex that is functionally isolated from the other TCP/IP stacks in the sysplex.

Procedure

Perform the following steps to partition a set of TCP/IP stacks in a sysplex into a subplex:

1. If TCP/IP coupling facility structures for SYSPLEXPORTS or sysplex-wide security associations (SWSA) are to be used, ensure that all the structures are defined to MVS and are specified in the active CFRM policy using their fully suffixed names. For example, if SYSPLEXPORTS is to be used with the base structure name of EZBEPORT, two VTAM subplexes are to be configured (11 and 12), and within each VTAM subplex, two TCP/IP subplexes are defined (21 and 22), ensure that the EZBEPORT1121, EZBEPORT1122, EZBEPORT1221, and EZBEPORT1222 structures are defined to MVS and specified in the active CFRM policy.
2. If the VTAM nodes are to be partitioned into subplexes, reconfigure each VTAM node in the VTAM subplex that will support the new TCP/IP subplex. For a description of how to set up VTAM subplexing, see z/OS Communications Server: SNA Network Implementation Guide.
3. As each VTAM is being reconfigured, for each TCP/IP stack that you want to place in a new subplex, perform the following steps:
   a) Stop the stack.
   b) Change the GLOBALCONFIG statement in the TCP/IP profile to include the XCFGRPID parameter with the group ID for the new subplex.
   c) If HiperSockets is enabled and there are TCP/IP stacks that are on the same CPC as this stack and that use the same HiperSockets CHPID, change the GLOBALCONFIG statement in the TCP/IP profile to include the IQDVLANID parameter with the selected VLAN ID for the new subplex.

      Tip: This step depends on the level of isolation you require for your subplexes within the sysplex. If you do not need to isolate the subplexes for HiperSockets connectivity (if, for example, you need only to restrict sysplex distributor distribution between subplexes), you do not need to perform this step.
   d) Start the TCP/IP stack, specifying the updated TCP/IP profile.
   e) Continue until all the TCP/IP stacks in the new subplex have been reconfigured.

Dynamic XCF

For most point-to-point links, a unique pair of IP addresses is needed for each stack within a sysplex. This requirement tends to use more IP addresses, and IP addresses can be saved by using dynamic XCF. Dynamic XCF creates a single IP address by which all other stacks in a sysplex can reach a stack. Dynamic XCF creates trusted, internal links to other stacks within a sysplex, generating dynamic definitions for TCP/IP stacks that are on another z/OS host in a sysplex or for TCP/IP stacks that are on the same z/OS host.

Dynamic XCF automatically generates the appropriate DEVICE, LINK, HOME, INTERFACE, BSDROUTINGPARMS, and BEGINROUTES definitions, and activates the devices to enable a stack to communicate with other stacks in the sysplex. Dynamic XCF devices and links, when activated, appear to the stack as though they had been defined in the TCP/IP profile, and can be displayed using standard commands.

Dynamic XCF support is available for both IPv4 and IPv6, and is enabled with the DYNAMICXCF parameter on the IPCONFIG or IPCONFIG6 statement, respectively. Though not specifically mentioned throughout this topic, unless otherwise noted, this information applies to IPv6 as well as IPv4.

Notes:

1. Dynamic XCF (non-IUTSAMEH and non-IQDIO) interface definitions to a particular LPAR are stopped and deleted when the last stack on a given LPAR is removed from the sysplex.
2. If you are using policy-based routing, by default dynamic XCF does not generate the corresponding policy-based routing RouteTable definitions. Determine whether it is appropriate for these routes to be added to each of your policy-based route tables. Use the DynamicXCFRoutes parameter on the RouteTable statement that defines a policy-based route table to indicate whether the routes should be added to that table. For information about the RouteTable statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

Getting started with dynamic XCF

The minimum requirements for TCP/IP stacks to use dynamic XCF differ based on whether same host or inter-host communication is being used. To generate definitions for two TCP/IP stacks that reside on different MVS hosts:

- Both MVS hosts must belong to the same sysplex.
- VTAM must have XCF communications enabled by specifying the XCFINIT=YES or XCFINIT=DEFINE start option, or by activating the VTAM XCF local SNA major node, ISTLSXCF. For details about configuration, see z/OS Communications Server: SNA Network Implementation Guide.
- DYNAMICXCF must be specified in the TCP/IP profile of each stack.

With this configuration, both same host and inter-host communication can be performed using dynamic XCF.

To generate definitions for two TCP/IP stacks that are on the same MVS host, you must specify DYNAMICXCF in the TCP/IP profile of each stack.

At initialization, each TCP/IP stack configured for XCF joins a well-known XCF group. When other stacks in the group discover the new stack, the definitions are created automatically, the links are activated, and the remote IP address for each link is added to the routing table. After the remote IP address has been added, IP traffic proceeds as usual.

In VTAM, you must activate the XCF local SNA major node. You can do this using the start option XCFINIT=YES or XCFINIT=DEFINE. If dynamically defined XCF definitions have been created for another VTAM in the sysplex that has since stopped and restarted with a different CPName, dynamic XCF recognizes this situation and automatically modifies existing definitions to accommodate the CPName change. If the XCF local SNA major node is inactive when TCP/IP is started and is not activated until after TCP/IP has finished initialization, TCP/IP does not generate any dynamic definitions for other TCP/IP hosts already started in the sysplex until either:

- A new TCP/IP host is detected

- A profile related operator command is issued (such as the VARY TCPIP,,OBEYFILE, VARY TCPIP,,START, or VARY TCPIP,,STOP commands)

Dynamic XCF for IPv4 addresses

To request dynamics for XCF or same host connections, enter the following definition in the IPCONFIG statement:

```
DYNAMICXCF IPAddress SubnetMask CostMetric
```

The internal definitions that are created depend on your configuration.

Scenario number 1

- TCP/IP detects another instance of TCP/IP on the same z/OS
- No device exists with the name IUTSAMEH
- No link exists with the name EZASAMEMVS

This scenario creates internal definitions equivalent to the following definitions:

DEVICE IUTSAMEH MPCPTP AUTORESTART
Device definition to obtain the most efficient stack-to-stack communications within the same MVS image.

LINK EZASAMEMVS MPCPTP IUTSAMEH
Link definition for the IUTSAMEH device.

HOME IPAddress EZASAMEMVS
Associates the IP address with the IUTSAMEH link.

BSDROUTINGPARMS EZASAMEMVS 65535 CostMetric SubnetMask DestIPAddress
Defines the link characteristics for EZASAMEMVS.

Note: The DestIPAddress is always 0.

START IUTSAMEH
Starts the IUTSAMEH device.

For details about XCF-related statements, see z/OS Communications Server: IP Configuration Reference. For information about changes to Netstat displays of dynamic XCF settings, see z/OS Communications Server: IP System Administrator's Commands.

Scenario number 2

- TCP/IP detects another instance of TCP/IP in the sysplex
- No device with the name of the CPName of the remote VTAM exists
- No HiperSockets connectivity between the two images exists
- No link exists with the name EZAXCF*nn*, where *nn* is the value of the MVS system symbol (SYSCLONE) for the MVS hosting the VTAM with the device name

This scenario creates internal definitions equivalent to the following definitions:

DEVICE CPName MPCPTP AUTORESTART
Device definition to communicate with TCP/IP stacks hosted by the remote VTAM.

LINK EZAXCF*nn* MPCPTP CPName
Link definition for the device, where *nn* is the SYSCLONE value for the remote VTAM and MVS.

HOME IPAddress EZAXCF*nn*
Associates the IP address with the dynamic XCF link.

BSDROUTINGPARMS EZAXCF*nn* 55296 CostMetric SubnetMask DestIPAddress
Defines the link characteristics for EZAXCF*nn*.

START CPName
Starts the specified device.

Notes:

1. If EZAXCF*nn* is already defined as a link name, or the CP name is already defined as a device name, then dynamic XCF definition EZAXCF*nn* and the CP name are not generated for discovery of another stack in the same sysplex.
2. The DestIPAddress is always zero.

For details about XCF-related statements, see z/OS Communications Server: IP Configuration Reference. For information about changes to Netstat displays of dynamic XCF settings, see z/OS Communications Server: IP System Administrator's Commands.

Scenario number 3

- TCP/IP detects another instance of TCP/IP in the sysplex
- The images are on the same CPC
- HiperSockets connectivity between the two images exists
- The host processor supports HiperSockets and z/OS Communications Server is properly configured
- No link exists with the name IQDIOLNK*nnnnnnnn* (where *nnnnnnnn* is the hexadecimal representation of the IP address specified on the IPCONFIG DYNAMICXCF statement)

This scenario creates internal definitions equivalent to the following definitions:

DEVICE IUTIQDIO MPCIPA AUTORESTART
Device definition to communicate with TCP/IP stacks hosted by the remote VTAM.

LINK IQDIOLNK*nnnnnnnn* IPAQIDIO IUTIQDIO
Link definition for the device, where *nnnnnnnn* is the hexadecimal representation of the IP address specified on the IPCONFIG DYNAMICXCF statement (that is, IPAddress).

HOME IPAddress IQDIOLNK*nnnnnnnn*
Associates the IP address with the dynamic XCF link.

BSDROUTINGPARMS IQDIOLNK*nnnnnnnn* 57344 CostMetric SubnetMask DestIPAddress
Defines the link characteristics for IQDIOLNK*nnnnnnnn*.

START IUTIQDIO
Starts the specified device.

Notes:

1. If IQDIOLNK*nnnnnnnn* is already defined as a link name, or IUTIQD*xx* (where *xx* is the CHPID of the HiperSockets LAN that is defined or specified by default on the VTAM start option IQDCHPID) is already defined as a device name, dynamic XCF IQDIOLNK*nnnnnnnn* and IUTIQDIO definitions are not generated for discovery of another stack in the same CPC.
2. The DestIPAddress is always zero.

For details about XCF-related statements, see z/OS Communications Server: IP Configuration Reference. For information about changes to Netstat displays of dynamic XCF settings, see z/OS Communications Server: IP System Administrator's Commands.

Dynamic XCF for IPv6 addresses

To request dynamics for XCF or same host connections, enter the following definition in the IPCONFIG6 statement:

```
DYNAMICXCF IP6Address
```

The internal definitions that are created depend on your configuration.

Scenario number 1

- TCP/IP detects another instance of TCP/IP on the same z/OS
- No device exists with the name IUTSAMEH

- No link exists with the name EZ6SAMEMVS

This scenario creates internal definitions equivalent to the following definitions:

INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR IP6Address
Interface definition for EZ6SAMEMVS.

START EZ6SAMEMVS
Starts the EZ6SAMEMVS interface.

For details about XCF-related statements, see z/OS Communications Server: IP Configuration Reference. For information about changes to Netstat displays of dynamic XCF settings, see z/OS Communications Server: IP System Administrator's Commands.

Scenario number 2

- TCP/IP detects another instance of TCP/IP in the sysplex
- No device with the name of the CPName of the remote VTAM exists
- No HiperSockets connectivity between the two images exists
- No link exists with the name EZ6XCF*nn*, where *nn* is the value of the MVS system symbol (SYSCLONE) for the MVS hosting the VTAM with the device name

This scenario creates internal definitions equivalent to the following definitions:

INTERFACE EZ6XCF*nn* DEFINE MPCPTP6 TRLENAME CPName IPADDR IP6Address
Interface definition, where *nn* is the SYSCLONE value for the remote VTAM and MVS.

START EZ6XCF*nn*
Starts the specified interface.

Notes:

1. If EZ6XCF*nn* is already defined as a link name, or the CP name is already defined as a device name, then dynamic XCF definition EZ6XCF*nn* and CP name are not generated for discovery of another stack in the same sysplex.
2. The DestIPAddress is always zero.

For details about XCF-related statements, see z/OS Communications Server: IP Configuration Reference. For information about changes to Netstat displays of dynamic XCF settings, see z/OS Communications Server: IP System Administrator's Commands.

IUTSAMEH

Communications Server provides internal links between TCP/IP stacks that are running within the same MVS image. This support is referred to as a Same Host (IUTSAMEH) link. If DYNAMICXCF is defined, TCP/IP always creates and activates a same host (IUTSAMEH) device and link (unless a static IUTSAMEH device is already defined) even if this is the only stack on the MVS image. When TCP/IP activates the IUTSAMEH device, VTAM dynamically builds the IUTSAMEH TRLE. The generated device name is IUTSAMEH, the generated link name is EZASAMEMVS (IPv4), and the generated interface name is EZ6SAMEMVS (IPv6). As other stacks are brought up within the same MVS image, a host route is created to each of these stacks across the same host link. It is recommended that users do not configure a static device for IUTSAMEH (allow TCP/IP to dynamically create the device and link). Communications Server also uses the IUTSAMEH link for Enterprise Extender support.

The IUTSAMEH device and link (IPv4) or interface (IPv6) do not become active and remain in SENT SETUP status until either another TCP/IP stack or Enterprise Extender connection is activated on this MVS image.

```
DEVNAME: IUTSAMEH        DEVTYPE: MPCPTP
DEVSTATUS: SENT SETUP
LNKNAME: EZASAMEMVS      LNKTYPE: MPCPTP      LNKSTATUS: NOT ACTIVE
⋮
INTFNAME: EZ6SAMEMVS     INTFTYPE: MPCPTP6     INTFSTATUS: NOT ACTIVE
```

In the case where IUTSAMEH is active solely because of Enterprise Extender and the IUTSAMEH is subsequently stopped, to restart the IUTSAMEH for Enterprise Extender, the XCA major node must be recycled. Otherwise, it might require manual reactivation of the Enterprise Extender LINEs and PUs, and manual redials.

XCF

When a subsequent stack within the sysplex is started that is not within the same MVS image, TCP/IP creates and activates an XCF device and link (unless an XCF device is already defined). The XCF links connect using the sysplex coupling facility or CTC links. A new device and link are created for every other VTAM node in the sysplex with at least one TCP/IP stack active on the same system with DYNAMICXCF specified. The generated device name is the CP name (for APPN) or SSCP name (for subarea-only nodes) of the remote VTAM. The generated link name is EZAXCF*nn* (IPv4), and the generated interface name is EZ6XCF*nn* (IPv6), where *nn* is the 2-character &SYSCLONE value. A host route across the XCF link is created when the XCF link is successfully activated.

Examples of definitions generated by dynamic XCF

This topic contains examples of definitions generated by dynamic XCF in both IPv4 and IPv6 environments. The notes following the examples pertain to both environments.

- IPv4 example 1:

  This configuration consists of two MVS systems (MVS1,MVS2) that are members of the same sysplex. Each MVS host has one TCP/IP stack (TCPIP1 and TCPIP2, respectively). From the syntax descriptions, the following information is needed to generate the dynamic definitions:

  – MVS sysclone value
  – VTAM CPName
  – Status of XCF in VTAM
  – The values specified on the IPCONFIG DYNAMICXCF keyword

  Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.1 255.255.255.248 3

MVS2:
Sysclone = B2 VTAM
Cpname = VTAM2
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.2 255.255.255.248 2
```

  After both TCPIP1 and TCPIP2 have been started, the following definitions will be generated.

  TCPIP1 will generate the equivalent of these definitions:.

```
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.1 EZAXCFB2
BSDROUTINGPARMS EZAXCFB2 55296 3 255.255.255.248 0
START VTAM2
```

  TCPIP2 will generate:

```
DEVICE VTAM1 MPCPTP AUTORESTART
LINK EZAXCFA1 MPCPTP VTAM1
HOME 9.1.1.2 EZAXCFA1
BSDROUTINGPARMS EZAXCFA1 55296 2 255.255.255.248 0
START VTAM1
```

When an XCF link becomes active, each TCPIP will generate a route to the other TCPIP over the XCF link. In this example, when the XCF link becomes active, TCPIP1 will generate a route to TCPIP2 over the XCF link and vice versa.

- IPv4 example 2:

  The configuration is the same as Example 1 except a second TCP/IP stack (TCPIP1A) was added to MVS1.

  Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.1 255.255.255.248 3
TCPIP1A: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.3 255.255.255.248 0

MVS2:
Sysclone = B2
VTAM Cpname = VTAM2
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.2 255.255.255.248 2
```

  After both TCPIP1 and TCPIP2 have been started, the following definitions will be generated, as in Example 1.

  TCPIP1 will generate the equivalent of these definitions:

```
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.1 EZAXCFB2
BSDROUTINGPARMS EZAXCFB2 55296 3 255.255.255.248 0
START VTAM2
```

  TCPIP2 will generate:

```
DEVICE VTAM1 MPCPTP AUTORESTART
LINK EZAXCFA1 MPCPTP VTAM1
HOME 9.1.1.2 EZAXCFA1
BSDROUTINGPARMS EZAXCFA1 55296 2 255.255.255.248 0
START VTAM1
```

  Now, TCPIP1A is started. TCPIP1 and TCPIP2 recognize that TCPIP1A has started. TCPIP1A will generate definitions for both TCPIP1 and TCPIP2. TCPIP1 will generate IUTSAMEH definitions for TCPIP1A. However, TCPIP2 does not need to generate and will not generate any new definitions except for routing information for TCPIP1A. New definitions do not need to be created because the DEVICE and LINK definitions are based on the discovery of a new VTAM node in the sysplex. (The DEVICE name is the VTAM CPName.)

  TCPIP1 will generate the equivalent of these definitions:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
HOME 9.1.1.1 EZASAMEMVS
BSDROUTINGPARMS EZASAMEMVS 65535 3 255.255.255.248 0
START IUTSAMEH
```

  When the IUTSAMEH connection becomes active, each TCPIP will generate a route to the other TCPIP over the IUTSAMEH connection.

  TCPIP2 does not generate anything.

  TCPIP1A will generate:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
DEVICE VTAM2 MPCPTP AUTORESTART
LINK EZAXCFB2 MPCPTP VTAM2
HOME 9.1.1.3 EZAXCFB2
HOME 9.1.1.3 EZASAMEMVS
BSDROUTINGPARMS EZAXCFB2 55296 0 255.255.255.248 0
```

```
BSDROUTINGPARMS EZASAMEMVS 65535 0 255.255.255.248 0
START IUTSAMEH
START VTAM2
```

When the IUTSAMEH connection becomes active, each TCPIP will generate a route to the other TCPIP over the IUTSAMEH connection.

- IPv4 example 3:

  To continue Example 2, add another MVS host (MVS3) with a VTAM node (VTAM3) with one TCP/IP stack (TCPIP3).

  ```
  MVS3:
  Sysclone = C3
  VTAM Cpname = VTAM3
  VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
  TCPIP3: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.3 255.255.255.248 0
  ```

  In this example, the previously active TCP/IP stacks will generate definitions for TCPIP3 because a new VTAM stack has become active in the sysplex. TCPIP3 will generate definitions for definitions for TCPIP1/TCPIP1A and TCPIP2.

  TCPIP1 will generate the equivalent of these definitions:

  ```
  DEVICE VTAM3 MPCPTP AUTORESTART
  LINK EZAXCFC3 MPCPTP VTAM3
  HOME 9.1.1.1 EZAXCFC3
  BSDROUTINGPARMS EZAXCFC3 55296 3 255.255.255.248 0
  START VTAM3
  ```

  TCPIP2 will generate:

  ```
  DEVICE VTAM3 MPCPTP AUTORESTART
  LINK EZAXCFC3 MPCPTP VTAM3
  HOME 9.1.1.2 EZAXCFC3
  BSDROUTINGPARMS EZAXCFC3 55296 2 255.255.255.248 0
  START VTAM3
  ```

  TCPIP1A will generate:

  ```
  DEVICE VTAM3 MPCPTP AUTORESTART
  LINK EZAXCFC3 MPCPTP VTAM3
  HOME 9.1.1.3 EZAXCFC3
  BSDROUTINGPARMS EZAXCFC3 55296 0 255.255.255.248 0
  START VTAM3
  ```

  TCPIP3 will generate:

  ```
  DEVICE VTAM1 MPCPTP AUTORESTART
  LINK EZAXCFA1 MPCPTP VTAM1
  DEVICE VTAM2 MPCPTP AUTORESTART
  LINK EZAXCFB2 MPCPTP VTAM2
  HOME 9.1.1.3 EZAXCFA1
  HOME 9.1.1.3 EZAXCFB2
  BSDROUTINGPARMS EZAXCFA1 55296 0 255.255.255.248 0
  BSDROUTINGPARMS EZAXCFB2 55296 0 255.255.255.248 0
  START VTAM1
  START VTAM2
  ```

- IPv4 example 4:

  This example illustrates how dynamic XCF can generate IUTSAMEH definitions without VTAM having its XCF enabled.

  ```
  MVS1:
  Sysclone = A1  (not used in this example)
  VTAM Cpname = VTAM1  (not used in this example)
  VTAM has XCFINIT=NO specified and has not activated the major node ISTLSXCF.
  TCPIP1: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.1 255.255.255.248 3
  TCPIP1A: PROFILE.TCPIP contains IPCONFIG DYNAMICXCF 9.1.1.3 255.255.255.248 0
  ```

  TCPIP1 will generate the equivalent of these definitions:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
HOME 9.1.1.1 EZASAMEMVS
BSDROUTINGPARMS EZASAMEMVS 65535 3 255.255.255.248 0
START IUTSAMEH
```

TCPIP1A will generate:

```
DEVICE IUTSAMEH MPCPTP AUTORESTART
LINK EZASAMEMVS MPCPTP IUTSAMEH
HOME 9.1.1.3 EZASAMEMVS
BSDROUTINGPARMS EZASAMEMVS 65535 0 255.255.255.248 0
START IUTSAMEH
```

- IPv6 example 1:

  This configuration consists of two MVS systems (MVS1,MVS2) that are members of the same sysplex. Each MVS host has one TCP/IP stack (TCPIP1 and TCPIP2, respectively). From the syntax descriptions, the following information is needed to generate the dynamic definitions:

  – MVS sysclone value
  – VTAM CPName
  – Status of XCF in VTAM
  – The values specified on the IPCONFIG6 DYNAMICXCF keyword

  Using the following user definitions:

```
 MVS1:
 Sysclone = A1
 VTAM Cpname = VTAM1
 VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
 TCPIP1: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f001
 MVS2:
 Sysclone = B2
 Cpname = VTAM2
 VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
 TCPIP2: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::222:f001
```

  After both TCPIP1 and TCPIP2 are started, the following definitions are generated.

  TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::111:f001
START EZ6XCFB2
```

  TCPIP2 generates:

```
 INTERFACE  EZ6XCFA1 DEFINE MPCPTP6 TRLENAME VTAM1 IPADDR 2001:0DB8::222:f001
 START EZ6XCFA1
```

  The INTERFACE statement combines the definitions of DEVICE, LINK, and HOME into a single statement for IPv6. When an XCF interface becomes active, each TCP/IP generates a route to the other TCP/IP over the XCF interface. In this example, when the XCF interface becomes active, TCPIP1 generates a route to TCPIP2 over the XCF interface, and TCPIP2 generates a route to TCPIP1 over the XCF interface.

- IPv6 example 2:

  The configuration is the same as Example 1 except a second TCP/IP stack (TCPIP1A) is added to MVS1.

  Using the following user definitions:

```
MVS1:
Sysclone = A1
VTAM Cpname = VTAM1
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP1: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f001
TCPIP1A: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f002
MVS2:
Sysclone = B2
VTAM Cpname = VTAM2
```

```
VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
TCPIP2: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::222:f001
```

After both TCPIP1 and TCPIP2 are started, the following definitions are generated, as in Example 1.

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::111:f001
START EZ6XCFB2
```

TCPIP2 generates:

```
INTERFACE  EZ6XCFA1 DEFINE MPCPTP6 TRLENAME  VTAM1 IPADDR 2001:0DB8::222:f001
START EZ6XCFA1
```

Now, TCPIP1A is started. TCPIP1 and TCPIP2 recognize that TCPIP1A has started. TCPIP1A generates definitions for both TCPIP1 and TCPIP2. TCPIP1 generates IUTSAMEH definitions for TCPIP1A. However, TCPIP2 does not need to generate and does not generate any new definitions, except for routing information, for TCPIP1A. New definitions do not need to be created because the INTERFACE definition is based on the discovery of a new VTAM node in the sysplex. (The TRLENAME is the VTAM CPName.)

TCPIP1 generates the equivalent of these definitions:

```
INTERFACE  EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME  IUTSAMEH IPADDR 2001:0DB8::111:f001
START EZ6SAMEMVS
```

When the IUTSAMEH connection becomes active, each TCP/IP generates a route to the other TCP/IP over the IUTSAMEH connection.

TCPIP2 does not generate anything.

TCPIP1A generates:

```
INTERFACE EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::111:f002
START EZ6XCFB2
INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR 2001:0DB8::111:f002
START EZ6SAMEMVS
```

When an XCF interface becomes active, each TCP/IP generates a route to the other TCP/IP over the XCF interface. In this example, when the XCF interface becomes active, TCPIP1A generates a route to TCPIP2 over the XCF interface, and TCPIP2 generates a route to TCPIP1A over the XCF interface. When the IUTSAMEH connection becomes active, each TCP/IP generates a route to the other TCP/IP over the IUTSAMEH connection.

- IPv6 example 3:

  To continue Example 2, add another MVS host (MVS3) with a VTAM node (VTAM3) with one TCP/IP stack (TCPIP3).

  ```
  MVS3:
  Sysclone = C3
  VTAM Cpname = VTAM3
  VTAM has either specified XCFINIT=YES, XCFINIT=DEFINE, or the major node ISTLSXCF is active
  TCPIP3: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::333:f001
  ```

  In this example, the previously active TCP/IP stacks generate definitions for TCPIP3 because a new VTAM stack has become active in the sysplex. TCPIP3 generates definitions for TCPIP1, TCPIP1A, and TCPIP2.

  TCPIP1 generates the equivalent of these definitions:

  ```
  INTERFACE  EZ6XCFC3 DEFINE MPCPTP6 TRLENAME VTAM3 IPADDR 2001:0DB8::111:f001
  START EZ6XCFC3
  ```

  TCPIP2 generates:

```
 INTERFACE  EZ6XCFC3 DEFINE MPCPTP6 TRLENAME VTAM3 IPADDR 2001:0DB8::222:f001
 START EZ6XCFC3
```

TCPIP1A generates:

```
INTERFACE  EZ6XCFC3 DEFINE MPCPTP6 TRLENAME VTAM3 IPADDR 2001:0DB8::111:f002
START EZ6XCFC3
```

TCPIP3 generates:

```
INTERFACE EZ6XCFA1 DEFINE MPCPTP6 TRLENAME VTAM1 IPADDR 2001:0DB8::333:f001
START EZ6XCFA1
INTERFACE EZ6XCFB2 DEFINE MPCPTP6 TRLENAME VTAM2 IPADDR 2001:0DB8::333:f001
START EZ6XCFB2
```

- IPv6 example 4:

  This example illustrates how dynamic XCF can generate IUTSAMEH definitions without VTAM having its XCF enabled.

```
MVS1:
Sysclone = A1  (not used in this example)
VTAM Cpname = VTAM1  (not used in this example)
VTAM has XCFINIT=NO specified and has not activated the major node ISTLSXCF.
TCPIP1: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f001
TCPIP1A: PROFILE.TCPIP contains IPCONFIG6 DYNAMICXCF 2001:0DB8::111:f002
```

  TCPIP1 generates the equivalent of these definitions:

```
INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR 2001:0DB8::111:f001
START EZ6SAMEMVS
```

  TCPIP1A generates:

```
 INTERFACE EZ6SAMEMVS DEFINE MPCPTP6 TRLENAME IUTSAMEH IPADDR 2001:0DB8::111:f002
 START EZ6SAMEMVS
```

Notes:

1. Because the interfaces generated by dynamic XCF use a single IP address, the output of the SIOCGIFCONF ioctl() contains multiple entries with the same IP address. If an application is using the SIOCGIFCONF output to issue bind() to all the entries returned, the application could receive EADDRINUSE on a bind() if there are multiple XCF devices defined by dynamic XCF in the list.
2. If you want to define a static route to a link which is generated by dynamic XCF, you must wait until the dynamic devices are started and then use the VARY TCPIP,,OBEYFILE command. The BEGINROUTES statement that refers to a dynamically defined link name or interface name must be in a separate data set from the data set that is used to define the dynamic devices (either the initial profile data set or the data set referenced by a VARY TCPIP,,OBEYFILE command).
3. Even though the HOME, BSDROUTINGPARMS and BEGINROUTES definitions are full replacement statements, the definitions generated by dynamic XCF will not replace any existing definitions. Likewise, user-defined HOME, BSDROUTINGPARMS and BEGINROUTES definitions will not affect existing or future definitions generated by dynamic XCF.
4. A mix of static and dynamic IPv4 and IPv6 definitions for a device is not allowed. For example, if a static IUTSAMEH IPv4 device and link is defined, an IPv6 dynamic definition for IUTSAMEH will not be created. If a static IUTSAMEH IPv6 interface is defined, an IPv4 dynamic definition for IUTSAMEH will not be created. The same logic also applies for XCF links; a mix of static and dynamic IPv4 and IPv6 definitions is not allowed for an XCF link.
5. The address specified on the IPCONFIG6 DYNAMICXCF statement cannot be an existing, statically defined, interface address. If a profile contains an IPv6 interface statement for an address that is also used in the IPCONFIG6 DYNAMICXCF statement, the IPCONFIG6 DYNAMICXCF statement is ignored.

Deleting dynamically defined XCF devices

You can delete dynamically defined XCF devices and links by first stopping the devices to be deleted and then issuing a VARY TCPIP,,OBEYFILE command that references a data set containing a DELETE LINK EZAXCF*nn* and DELETE DEVICE. Because the HOME statement processing does not affect dynamically defined XCF HOME list entries, the HOME *xx.xx.xx.xx* EZAXCF*nn* entry is automatically deleted by DELETE LINK.

For IPv6, you can delete dynamically defined XCF interfaces by first stopping the interface previously defined in an INTERFACE statement and then issuing a VARY TCPIP,,OBEYFILE command that references a data set containing an INTERFACE *interfacename* DELETE statement.

HiperSockets

HiperSockets is a System z hardware feature that provides high performance internal communications between LPARs within the same central processor complex (CPC) without the use of any additional or external hardware equipment (for example, channel adapters, LANs, and so on).

If the host processor supports HiperSockets and Communications Server is properly configured, Communications Server will attempt to create XCF connectivity between LPARs in the same CPC using a HiperSockets link or interface. When the HiperSockets link or interface cannot be activated, TCP/IP creates a normal XCF link or interface. However, if the HiperSockets link or interface is successfully activated, but is then later stopped, TCP/IP does not create an XCF link or interface.

The HiperSockets for DYNAMICXCF device and link or interface are dynamically built and the device or interface is started during TCP/IP DYNAMICXCF stack initialization. The HiperSockets for DYNAMICXCF device and link and interface cannot be configured. The generated device name is IUTIQDIO. The generated IPv4 link name is IQDIOLNK*nnnnnnnn*, where *nnnnnnnn* is the character representation of the hexadecimal version of the DYNAMICXCF IP address. The generated IPv6 interface name is IQDIOINTF6. In general, where an XCF link or interface would normally have been used (for intra-CPC), a HiperSockets link or interface will be used.

Similar to IUTSAMEH, VTAM will dynamically build the TRLE for IUTIQDIO when the IUTIQDIO device is started. The TRLE statement is not configured (defined) by the user.

Although HiperSockets for DYNAMICXCF is not configured with TCP/IP DEVICE and LINK or INTERFACE statements, and the TRLE is not defined by the user, the following steps must be taken to define the HiperSockets subchannel devices and IQD CHPID:

1. Using HCD or IOCP, the system administrator must define (create the IOCDS) the HiperSockets IQD CHPID (channel path ID) and subchannel devices to the applicable LPARs. To dynamically build the HiperSockets TRLE, VTAM requires a minimum of 3 subchannel devices configured with each IQD CHPID within HCD. The maximum number of subchannel devices that VTAM will use (associate with each TRLE or MPC group) is 10. For additional details regarding configuring the HiperSockets subchannel devices and IQD CHPID, see z/OS HCD Planning and Appendix D, “Using HCD,” on page 1339.
2. When more than one IQD CHPID is configured to a specific LPAR, VTAM start option IQDCHPID must be used to specify which specific IQD CHPID this LPAR should use. The VTAM start option controls which IQD CHPID (and related subchannel devices) VTAM selects to include in the HiperSockets MPC group (IUTIQDIO) when it is dynamically built for HiperSockets for DYNAMICXCF connectivity. Start option IQDCHPID controls the VTAM IQD CHPID selection for the HiperSockets for DYNAMICXCF device IUTIQDIO (MPC group) only. It does not control IQD CHPID selection for a user defined HiperSockets device (MPCIPA). However, a user defined HiperSockets device (IQD CHPID) cannot use (conflict with) the same IQD CHPID that the HiperSockets for DYNAMICXCF device is currently using.

   For example, if IQD CHPID 'FE'x is in use by DYNAMICXCF because of one of the following conditions, then an attempt to configure and start a user defined HiperSockets device IUTIQDFE is not allowed (IQD CHPIDs conflict):

   a. VTAM start option IQDCHPID=FE is specified

b. VTAM start option IQDCHPID=ANY is specified, but the HiperSockets for DYNAMICXCF device IUTIQDIO is using the 'FE' CHPID

This option can also be modified with a VTAM modify command. In most cases, the default setting will be sufficient. For additional details regarding this start option, see z/OS Communications Server: SNA Resource Definition Reference.

For additional details regarding HiperSockets, see "HiperSockets concepts and connectivity" on page 75.

Network interfaces monitoring

Sysplex distributor and other dynamic VIPA (DVIPA) functions depend on an available network path to the TCP/IP stack that owns and advertises a dynamic VIPA. If connectivity is disrupted, clients might not be able to access the applications represented by the DVIPA, impacting their operations. Dynamic XCF interfaces can be configured to provide for a backup network path for DVIPAs. If the dynamic XCF interfaces fail for a given system, the sysplex autonomics function might automatically initiate a recovery action, resulting in the local TCP/IP stack leaving the sysplex and enabling other TCP/IP stacks to assume ownership of the DVIPAs.

However, there are some configurations in which it is not optimal to configure dynamic XCF interfaces as eligible backup network paths for TCP/IP stacks that own and advertise DVIPAs. In these configurations, incoming network traffic for the DVIPAs is expected to arrive over one or more external network interfaces. If these external interfaces fail, or the networks they are attached to experience a failure, the DVIPAs owned by the local TCP/IP stack can become isolated. Client traffic destined for these DVIPAs cannot reach the local TCP/IP stack; the DVIPA is unreachable.

The network interfaces monitoring function enhances the sysplex autonomics function by enabling you to specify key network interfaces that should be monitored by TCP/IP stacks. If a failure occurs on all specified interfaces, sysplex autonomics recovery can be triggered so other TCP/IP stacks in the sysplex can take over responsibilities for the DVIPAs owned by the local stack.

This level of monitoring is useful not only for TCP/IP stacks that currently own and advertise distributed and non-distributed DVIPAs, but also for TCP/IP stacks that are eligible backup stacks for these DVIPAs. This can help ensure that any backup TCP/IP stacks that are experiencing network connectivity problems do not attempt to perform DVIPA takeover activities if the primary DVIPA owner is stopped or fails. By being proactive and detecting these network connectivity problems, the backup stack can remove itself from the sysplex, enabling other healthy backup stacks to perform the takeover activities. While this monitoring function can also be enabled on TCP/IP stacks that act only as targets for distributed DVIPAs, the benefits for this configuration are minimal, as sysplex distributor already automatically monitors the ability of target TCP/IP stacks to communicate with distributed DVIPA clients.

The following functions are provided by the network interfaces monitoring function:

- You can identify which links and interfaces are critical for inbound network connectivity for a distributed DVIPA or DVIPA-owning TCP/IP stack. Specify which links and interfaces are to be monitored by sysplex autonomics using the MONSYSPLEX parameter on the LINK and INTERFACE statements.
- You have the option of enabling the TCP/IP stack to monitor the status of these interfaces through the MONINTERFACE keyword on the GLOBALCONFIG SYSPLEXMONITOR statement. If the MONINTERFACE keyword is specified, by default the TCP/IP stack also monitors the presence of dynamic routes over these interfaces. To override monitoring of dynamic routes, specify the MONINTERFACE NODYNROUTE keyword on the GLOBALCONFIG SYSPLEXMONITOR statement. The monitoring of dynamic routes enables TCP/IP to extend its monitoring beyond the status of the actual network interface, to also include the status of any routers that are connected to the local area network to which these interfaces are attached. Because a dynamic routing protocol is usually required for DVIPAs, by checking for the presence of dynamic routes, TCP/IP can determine whether the local OMPROUTE daemon is successfully communicating with these routers using dynamic routing protocols. If these network interfaces and dynamic routes become unavailable or experience network connectivity failures, the sysplex autonomics function can automatically initiate a recovery action.

If network interfaces monitoring is enabled, you must be aware of the following scenarios where the monitoring function can create problems:

- Monitored interfaces are deleted

  To delete interfaces, the devices or interfaces must first be stopped. When the devices or interfaces are stopped, the interface becomes inactive. If the TCP/IP stack detects that all monitored interfaces are inactive as a result of stopping devices or interfaces, the TCP/IP stack can issue messages regarding the problem and possibly trigger a recovery action. You can disable monitoring of these interfaces before stopping the devices or interfaces, using the VARY TCPIP,,OBEYFILE command and specifying the NOMONSYSPLEX keyword for the LINK or INTERFACE statement. For more information about DEVICE and LINK statements and INTERFACE statements, see z/OS Communications Server: IP Configuration Reference.
- The first hop routers are brought down for maintenance

  If the MONINTERFACE keyword or MONINTERFACE DYNROUTE option is configured on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and the first hop routers are brought down for maintenance, an inadvertent recovery action might be triggered. You can temporarily disable the monitoring of dynamic routes using the VARY TCPIP,,OBEYFILE command and specifying MONINTERFACE NODYNROUTE on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement.
- Preferred routing interface is not monitored

  If you use multiple interfaces and at least one interface is preferred from a routing perspective over the other interfaces, ensure that the preferred interfaces are monitored (the MONSYSPLEX keyword is specified for the LINK or INTERFACE statement). If the preferred interfaces are not monitored, an inadvertent recovery action might be initiated.

For more information about sysplex autonomics, see "Sysplex problem detection and recovery" on page 452.

Sysplex problem detection and recovery

Each sysplex member monitors itself and can automatically leave the sysplex group if it determines that it is no longer able to function correctly as a router, target, backup, or owner of a DVIPA. Through a variety of methods, it monitors:

- Internal resources and conditions, such as timely execution of sysplex functions, available private storage, and common storage
- External resources, such as availability of VTAM, or OMPROUTE if it is being used, or IPsec infrastructure (such as IKED) if IPsec is being used.

As long as the TCP/IP stack is a member of a TCP/IP sysplex group, the sysplex monitor gets control periodically. The time period is determined from the TIMERSECS value specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement in PROFILE.TCPIP. The default TIMERSECS value is 60 seconds.

After a problem is detected, further actions depend on whether RECOVERY or NORECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. NORECOVERY is the default value.

Joining the TCP/IP sysplex group can be delayed until certain network routing and connectivity availability conditions are met. Those conditions can include any combination of the following conditions:

- OMPROUTE is active
- Monitored interfaces are active
- Dynamic routes over those monitored interfaces are present
- IPsec components are active and operational

The first condition is activated using the DELAYJOIN option on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. The next two conditions are activated using the MONINTERFACE or MONINTERFACE DYNROUTE option on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. The last condition is activated by the DELAYJOINIPSEC option on the SYSPLEXMONITOR parameter on the GLOBALCONFIG statement. No sysplex-related definitions within the TCP/IP profile (that is, VIPADYNAMIC and DYNAMICXCF statements) are processed until the sysplex group is joined.

Tips:

- To determine whether the stack has joined a sysplex group, issue the DISPLAY TCPIP,,SYSPLEX,GROUP command.
- If you have procedures specified on the AUTOLOG profile statement that bind to a dynamic VIPA, specify the optional DELAYSTART parameter with the optional DVIPA subparameter for these procedures on the AUTOLOG statement. With the DELAYSTART DVIPA subparameter, the procedures will not start until the TCP/IP stack has joined the sysplex group and has processed the dynamic VIPA configuration.

During a planned or unplanned outage, the DVIPAs and distributed DVIPAs for a TCP/IP stack are taken over by backup TCP/IP stacks. When the primary TCP/IP stack is restarted, the DVIPAs and distributed DVIPAs are taken back from the backup TCP/IP stacks.

- If dynamic routing is used to advertise routes to these DVIPAs, and specified network routing and connectivity availability conditions are not met on the primary TCP/IP stack, existing connections to these DVIPAs might be reset and new connect requests to these DVIPAs might fail. By using the GLOBALCONFIG SYSPLEXMONITOR DELAYJOIN and MONINTERFACE DYNROUTE configuration statements in the TCP/IP profile on the primary TCP/IP stack, it is possible to delay taking back the DVIPAs and distributed DVIPAs until specified network routing and connectivity availability conditions are met. New and existing connections continue to be serviced by the backup TCP/IP stacks until OMPROUTE is active and monitored interfaces and dynamic routes over those monitored interfaces are present on the primary TCP/IP stack.
- When IPsec infrastructure is needed to establish IPsec protection for traffic using these DVIPAs, and all the necessary IPsec components are not yet operational on the primary TCP/IP stack, existing connections to these DVIPAs might be reset and new connect requests to these DVIPAs might fail. By using the GLOBALCONFIG SYSPLEXMONITOR DELAYJOINIPSEC configuration statement in the TCP/IP profile on the primary TCP/IP stack, it is possible to delay taking back the DVIPAs and distributed DVIPAs until the IPsec infrastructure is operational. New and existing connections continue to be serviced by the backup TCP/IP stacks until the IPsec infrastructure is operational on the primary TCP/IP stack.

For more information about the GLOBALCONFIG statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

Problem detection

If this stack is not the only member of the TCP/IP sysplex group and it is advertising DVIPAs (owns a DVIPA or is the primary routing node for a Distributed DVIPA), the following resource checks are made by the monitor:

- VTAM address space availability

  If the VTAM address space is not currently active and the elapsed time since VTAM was last detected as active has exceeded the TIMERSECS value, message EZZ9671E is issued.

- Route availability

  If there are no routes available to all partners, given the configured method chosen for forwarding data (VIPAROUTE statement or dynamic XCF interfaces), there are at least two other MVS systems in the sysplex, and the elapsed time since an active route was detected has exceeded the TIMERSECS value, message EZZ9673E or EZD1172E is issued.

If OMPROUTE was successfully initialized, it periodically sends a heartbeat to the stack, so the monitor can always make the following resource checks:

- If the elapsed time since a heartbeat was received has exceeded half of the TIMERSECS value, message EZZ9672E is issued as a warning and no other actions occur.
- If the stack is not the only member of a TCP/IP sysplex group, is advertising DVIPAs, and the elapsed time since a heartbeat was received has exceeded the TIMERSECS value, message EZZ9678E is issued.

Ensure that the WLM policy for the OMPROUTE address space receives sufficient priority in relationship to other work managed by WLM on the system, so that OMPROUTE receives the CPU cycles necessary for this task. For more information on this topic, see Sysplex autonomics.

If the network monitoring function is enabled, at least one monitored interface is configured, the TCP/IP stack is not the only member of this sysplex group, and the TCP/IP stack participates in sysplex distribution (as a distributor or target) or acts as an owner or a backup for DVIPAs, the following network connectivity checks are made by the monitor:

- Critical interfaces are active

  If the MONINTERFACE option is specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, all monitored interfaces become inactive, and the elapsed time since at least one active monitored interface was detected has exceeded the TIMERSECS value, message EZD1209E is issued.
- Presence of dynamic routes available over critical interfaces

  If the MONINTERFACE or MONINTERFACE DYNROUTE option is specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, there are no available dynamic routes found over any of the monitored interfaces, and the elapsed time since at least one dynamic route over the monitored interfaces was detected has exceeded the TIMERSECS value, message EZD1210E is issued.

If the IPsec monitoring function is enabled (SYSPLEXMONITOR DELAYJOINIPSEC MONIPSEC on the GLOBALCONFIG statement) and sysplex-wide security associations (SWSA) are enabled with the DVIPSEC parameter on the IPSEC statement, the IPsec infrastructure is monitored. The IKE daemon periodically sends a heartbeat to the stack. The monitor checks if the elapsed time since a heartbeat was received has exceeded the TIMERSECS value. If so, message EZD1977E is issued.

If this stack is advertising DVIPAs or is a sysplex distributor target, the following checks are made:

- The monitor checks for the availability of CSM storage. If CSM storage is continuously critical for greater than the TIMERSECS value, message EZZ9679E is issued. If CSM storage is continuously constrained for greater than three times the TIMERSECS value, message EZD1974E is issued.
- If TCP/IP ECSA storage limits are defined on the ECSALIMIT parameter of the GLOBALCONFIG statement, monitoring of this storage is similar to CSM monitoring. If TCP/IP ECSA storage is continuously critical for a time greater than the TIMERSECS value, message EZD1187E is issued. If there are no TCP/IP ECSA storage limits defined on the GLOBALCONFIG statement and a storage request cannot be satisfied, message EZD1170E is issued. These messages indicate that there is a TCP/IP problem with ECSA storage.
- If TCP/IP private storage limits are defined on the POOLLIMIT parameter of the GLOBALCONFIG statement, monitoring of this storage is similar to CSM monitoring. If TCP/IP private storage is continuously critical for a time greater than the TIMERSECS value, message EZD1187E is issued. If there are no TCP/IP private storage limits defined on the GLOBALCONFIG statement and a storage request cannot be satisfied, message EZD1170E is issued. These messages indicate that there is a TCP/IP problem with private storage.

Note, however, that if all DVIPAs on this stack have a status of MOVING and the stack is not a DVIPA target, these checks are not made.

In addition to the sysplex monitor, when the stack joins the TCP/IP sysplex group it requests that XCF monitor the TCP/IP sysplex component on the local stack. The XCF component performs this function by monitoring a status field updated by the TCP/IP sysplex component. If XCF detects that the sysplex functions have not been responsive for the TIMERSECS value, it schedules a TCP/IP routine that issues message EZZ9674E.

If TCP/IP encounters a nonrecoverable sysplex error, it issues the eventual action message EZZ9670E.

If five TCP/IP abends occur in less than 1 minute, eventual action message EZD1973E is issued.

All of these messages are eventual action messages. If the detected problem condition is corrected (for example, VTAM is started), the eventual action message is cleared. For more information about these eventual action messages, see z/OS Communications Server: IP Messages Volume 2 (EZB, EZD) and z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM).

Recovery

Recovery actions occur if RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. If the default setting, SYSPLEXMONITOR NORECOVERY, is active, other than issuing the message, no further actions occur if the problem is not corrected.

The VARY TCPIP,,SYSPLEX,LEAVEGROUP command can be used to manually force the sysplex member to leave the TCP/IP sysplex group. As a stack leaves the TCP/IP sysplex group, message EZZ9670E is cleared, as well as message EZD1170E. All other outstanding eventual action messages are cleared when the condition is cleared (for example, starting VTAM). For information on the VARY TCPIP,,SYSPLEX command, see z/OS Communications Server: IP System Administrator's Commands.

If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement in the TCP/IP profile, and this stack is not the only member of the TCP/IP sysplex group, the stack leaves the TCP/IP sysplex group when one of the messages is issued. The one exception to this is EZZ9672E, which is issued only as an OMPROUTE warning message. No actions occur unless the corresponding EZZ9678E OMPROUTE message is subsequently issued.

To determine whether the stack is currently joined to a TCP/IP sysplex group, issue the DISPLAY TCPIP,,SYSPLEX,GROUP command. If the stack is not currently joined to a TCP/IP sysplex group, this command displays the following message:

```
EZZ8269I tcpstackname mvsname NOT A MEMBER OF A SYSPLEX GROUP
```

If the stack is currently joined, the name of the TCP/IP XCF group is displayed.

From any member of the sysplex, use the D XCF,GROUP,*groupname* command to see the systems currently in the sysplex group, where *groupname* is EZBTCPCS, or if subplexing is being used, EZBT*vvtt*, where *vv* is the VTAM XCF group ID suffix and *tt* is the TCP group ID suffix..

If the RECOVERY option is specified and a TCP/IP stack initiates an automated recovery action by leaving the TCP/IP sysplex group, all local DVIPAs are deactivated and all the VIPADYNAMIC block definitions are saved. Any applications bound to dynamically created DVIPAs (VIPARANGE or MODDVIPA) will receive an asynchronous error, `EUNATCH (3448) - the protocol required to support the address family is unavailable`.

If internal problems prevent the removal of these resources, eventual action message EZZ9675E is issued, and restarting the stack is necessary to be able to become part of the TCP/IP sysplex group. If all DVIPAs are successfully removed, eventual action message EZZ9676E is issued, indicating that sysplex problem detection cleanup has succeeded. There are two ways for the stack to rejoin the sysplex group and clear this message after a successful cleanup has occurred:

- If AUTOREJOIN was configured on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, the stack automatically rejoins the group and reprocesses its saved VIPADYNAMIC configuration when all detected problems have been relieved. The AUTOREJOIN option is the recommended setting when the RECOVERY option is configured.
- Issue the VARY TCPIP,,SYSPLEX,JOINGROUP command to cause the stack to rejoin the group, and reprocess its VIPADYNAMIC configuration.

Recovery is the preferred method of operation, because this allows other members of the TCP/IP sysplex to automatically take over the functions of a member with no actions needed by an operator. IBM Health Checker for z/OS can be used to check whether the RECOVERY parameter has been specified when the IPCONFIG DYNAMICXCF or the IPCONFIG6 DYNAMICXCF statement has been specified. For more details about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide.

There are, however, some environments and scenarios where this automated recovery action might not be desirable and perhaps should not be enabled:

- DVIPAs or Distributed DVIPAs are defined, but no backup TCP/IP stacks are identified or no provisions are made to move the DVIPAs in cases of failure.

  The basic premise of the automated recovery actions is that one or more other TCP/IP stacks in the system can pick up ownership responsibilities for any DVIPAs owned by the failing TCP/IP stack. If

this is not the case, it is suggested that you carefully evaluate the benefits of designating backup TCP/IP stacks and implement a configuration that includes backup capabilities. If this is not possible or desirable, RECOVERY should not be specified. If RECOVERY is not specified, automated recovery actions are disabled by default.

For example, one such configuration is if you are using only DVIPAs that are always bound to a specific TCP/IP stack (that is, in lieu of static VIPAs). In this scenario, because there is no possibility of having ownership of these DVIPAs transferred automatically, there is no value in triggering the automated recovery action and you should consider not enabling the automated recovery function or using static VIPAs (because static VIPAs are not affected by the automated recovery actions).

- Test environments where individual system images have very limited resources (CPU, storage, and so on).

  This can include environments where you are running z/OS as a second level guest under z/VM, or in LPARs with shared processors and very limited resources. Not enabling the automated recovery actions in these environments can help prevent unwanted recovery actions that are triggered by false positive conditions, such as scenarios where artificial severe resource shortages are detected.

- Environments where VTAM, OMPROUTE, IKED or NSSD are stopped for intervals longer than the TIMERSECS value specified on the SYSPLEXMONITOR parameter.

  If your current operations procedures have provisions for stopping VTAM, OMPROUTE, IKED or NSSD for extended periods of time, you should consider disabling and re-enabling the automated recovery processing around the periods of time where you stop and restart these components. This can be accomplished using the VARY TCPIP,,OBEYFILE command.

  An alternative solution could be to increase the TIMERSECS value to accommodate the longest period of time you would expect VTAM, OMPROUTE, IKED or NSSD to be inactive during normal operating procedures. One potential drawback of this approach is that the monitoring of other conditions and triggering of automatic recovery functions is less responsive.

Setting TIMERSECS

As described, the responsiveness of the self-monitoring and automated recovery actions is governed by the TIMERSECS value that is specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement. The default value is 60 seconds. Evaluate whether this default value is appropriate in your environment and, if not, specify a value that is better suited to your environment. The following considerations are some of the considerations to evaluate in selecting an appropriate value:

- Specifying a smaller TIMERSECS value increases the responsiveness of the self-monitoring functions.

  The monitor periodically performs its checks for problem conditions, every 60 seconds or every quarter of the TIMERSECS interval (TIMERSECS/4), whichever is less. Therefore, when a smaller TIMERSECS value is specified, operator message alerts are issued sooner if a problem condition is detected. Automated recovery actions are also triggered sooner if the RECOVERY option is in effect.

  While this might be desirable, care should be taken to not specify a timer interval value that is small enough to trigger recovery actions for conditions that are transient in nature. For example, the value specified should probably be larger than the spin loop timeout interval for the system (SPINTIME parameter in the EXSPAT*xx* parmlib member). Otherwise, a spin loop timeout condition that might be recoverable could trigger an unnecessary TCP/IP sysplex recovery action, as a result of the TCP/IP sysplex monitor not being able to be dispatched temporarily before any spin loop timeout recovery actions can take place. A good rule of thumb is to specify a TIMERSECS value that is at least two times larger than the spin loop timeout interval in effect for the system.

- Specifying a longer TIMERSECS value has the opposite effect, and the self-monitoring and automated recovery actions are less responsive.

  In addition, a longer TIMERSECS value can cause disruptions to existing TCP connections when problem conditions are detected, even if automated recovery actions are initiated. For example, if the problem condition persists for a long enough interval and it impacts the delivery of data on existing TCP connections, the remote TCP/IP hosts might terminate these connections before a recovery action is initiated.

- Review your current settings for the sysplex failure detection interval (INTERVAL keyword in COUPLE*xx*) and the ISOLATETIME value specified in your SFM policy, if defined. These values indicate how responsive the XCF component should be in removing systems from the sysplex when a status update missing condition occurs. As a result, these settings can provide a reasonable reference point for the setting of the TIMERSECS value. For more information on these intervals, see z/OS MVS Setting Up a Sysplex.

Summary of problems monitored and actions taken

When the sysplex monitor detects a problem, an operator message is issued. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement in the TCP/IP profile, the stack leaves the sysplex group, deletes all its dynamic VIPAs, and deactivates all its VIPADYNAMIC definitions. When the detected problem is relieved, the operator message is deleted. If AUTOREJOIN was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, the stack might or might not rejoin the sysplex group, depending on the severity of the problem.

Table 25 on page 457 summarizes the various problems that are monitored and the specific actions that are taken for each detected problem.

Table 25. Sysplex problem monitoring

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
VTAM address space is not active	DVIPAs advertised and VTAM address space has been inactive greater than TIMERSECS interval	1. Issue EZZ9671E, VTAM inactive for at least TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when VTAM is started.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
Dynamic or static XCF route, or VIPAROUTE not active	DVIPAs advertised, acting as forwarder for a DVIPA, two other MVS systems in the sysplex, and either XCF routes (when VIPAROUTE is not specified) or all IP routes (when VIPAROUTE is enabled) to all sysplex partners are not available for TIMERSECS interval	1. If VIPAROUTE is not enabled, issue message EZZ9673E , dynamic XCF connectivity to all partners not available for at least TIMERSECS interval. 2. If VIPAROUTE is enabled, issue message EZD1172E , dynamic XCF connectivity and IP routes to all partners are not available for at least TIMERSECS interval. 3. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 4. Delete the operator message when any VIPAROUTE or XCF route becomes active.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
Network connectivity – critical interfaces are not active	• MONINTERFACE is specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement and at least one interface is configured with the MONSYSPLEX keyword. • The TCP/IP stack participates in sysplex distribution (as a distributor or target) or acts as an owner or a backup for DVIPAs. • Another TCP/IP member exists in the sysplex. • All monitored interfaces become inactive for the TIMERSECS interval.	1. Issue message EZD1209E when all monitored interfaces become inactive for at least the TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, leave the sysplex group and delete all VIPADYNAMIC block definitions. 3. Delete the operator message when any monitored interface becomes active.	

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
Network connectivity – dynamic routes are not present over critical interfaces	• MONINTERFACE or MONINTERFACE DYNROUTE is specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and at least one interface is configured with the MONSYSPLEX keyword. • The TCP/IP stack participates in sysplex distribution (as a distributor or target) or acts as an owner or a backup for DVIPAs. • Another TCP/IP member exists in the sysplex. • No dynamic routes are found over critical interfaces for the TIMERSECS interval.	1. Issue message EZD1210E when no dynamic routes over critical interfaces are found for at least the TIMERSECS interval 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, leave the sysplex group and delete all VIPADYNAMIC block definitions. 3. Delete the operator message when any dynamic route over a critical interface is found.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
OMPROUTE heartbeats are no longer being received	First heartbeat received, and no heartbeat received for half of the TIMERSECS interval	1. Issue message EZZ9672E, the OMPROUTE not responsive warning message. 2. Delete the operator message when heartbeat received, or when message EZZ9678E is issued due to OMPROUTE not responding for at least TIMERSECS interval.	Not applicable

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
OMPROUTE heartbeats are no longer being received	First heartbeat received, DVIPAs advertised, and no heartbeat received for TIMERSECS interval	1. Issue message EZZ9678E, OMPROUTE not responsive for at least TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. If a last heartbeat was not received (OMPROUTE did not terminate properly), abend and dump the TCP/IP and OMPROUTE address space. 3. Delete the operator message when heartbeat received.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
TCP/IP private storage allocation failures, when POOLLIMIT is not coded on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP private storage allocation fails	1. Issue message EZD1170E, TCP/IP private storage allocation failed message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP private allocation requests are successful for a full monitor interval (1/4 of the TIMERSECS interval or 1 minute, whichever is less), or delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
TCP/IP private critical storage limits reached, when POOLLIMIT values are specified on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP/IP private pool is critical for TIMERSECS interval. If TCP/IP private storage allocation request from MVS fails before the TCP/IP private pool is critical for TIMERSECS interval, the problem is treated as if no POOLLIMIT values were specified.	1. Issue message EZD1187E, TCP/IP private storage is critical. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP private pool storage exits the critical state.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
TCP/IP ECSA critical storage limits reached, when TCP/IP ECSA storage limits are specified on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP/IP ECSA is critical for TIMERSECS interval. If TCP/IP ECSA storage allocation request from MVS fails before TCP/IP is critical for TIMERSECS interval, the problem is treated as if no TCP/IP limits were specified.	1. Issue message EZD1187E, TCP/IP ECSA storage is critical. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP ECSA critical state is exited.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
TCP/IP ECSA storage allocation failures, when no TCP/IP ECSA storage limits are specified on the GLOBALCONFIG statement	DVIPAs advertised or this is a DVIPA target, and TCP/IP ECSA storage allocation request fails	1. Issue message EZD1170E, TCP/IP ECSA storage allocation failed message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when TCP/IP ECSA allocation requests are successful for a full monitor interval (1/4 of the TIMERSECS interval or 1 minute, whichever is less), or delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.
CSM storage critical	DVIPAs advertised or this is a DVIPA target, and CSM critical for greater than TIMERSECS interval	1. Issue message EZZ9679E, CSM critical message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when CSM critical state exited.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
CSM storage constrained	DVIPAs advertised or this is a DVIPA target, and CSM constrained for greater than three times the TIMERSECS interval	1. Issue message EZD1974E, CSM constrained message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when CSM constrained state exited.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
XCF status is not being updated	Status not updated for TIMERSECS interval	1. Issue message EZZ9674E, sysplex processing not responsive message. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when the monitor is able to run, or delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
Abend in sysplex code		1. Issue message EZZ9670E, sysplex processing encountered nonrecoverable error. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.
Five abends in less than 1 minute		1. Issue message EZD1973E, multiple nonrecoverable errors are adversely affecting sysplex processing. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when the stack leaves the TCP/IP sysplex group.	Stack does not rejoin the sysplex group. Use of the VARY TCPIP,,SYSPLEX,JOINGROUP command is required to rejoin the sysplex and restore the saved VIPADYNAMIC configuration.

Table 25. Sysplex problem monitoring (continued)

Type of check	When problem is monitored	Action taken when problem detected	Action taken if AUTOREJOIN coded and problem is cleared
IKED heartbeats are no longer being received (see Sysplex autonomics for IPsec for more information)	MONIPSEC is specified on the SYSPLEXMONITOR DELAYJOINIPSEC parameter of the GLOBALCONFIG statement, DVIPSEC parameter specified on IPSEC statement and no heartbeat received for TIMERSECS interval	1. Issue message EZD1977E, IPsec infrastructure not responsive for at least TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when heartbeat received.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration
IPsec policy filters are no longer in effect in the TCP/IP stack	MONIPSEC is specified on the SYSPLEXMONITOR DELAYJOINIPSEC parameter of the GLOBALCONFIG statement, DVIPSEC parameter specified on the IPSEC statement and default IPsec filters in effect	1. Issue message EZD1979E, Default IPsec filters have been in effect for at least TIMERSECS interval. 2. If RECOVERY was specified on the SYSPLEXMONITOR parameter of the GLOBALCONFIG statement, and this TCP/IP stack is not the only member of the TCP/IP sysplex group, leave the sysplex group, deactivate all DVIPAs, and save all VIPADYNAMIC block definitions. 3. Delete the operator message when IPsec policy filters in effect.	Stack rejoins the sysplex group and reprocesses the saved VIPADYNAMIC configuration

Sysplex autonomics for IPsec

Sysplex autonomics for IPsec provides:

- a mechanism for a TCP/IP stack to delay joining or rejoining the sysplex group until the IPsec infrastructure is active
- a mechanism for the TCP/IP stack to detect a problem in the IPsec infrastructure and optionally leave the sysplex group until the problem is resolved

You enable sysplex autonomics for IPsec by specifying the DELAYJOINIPSEC and MONIPSEC subparameters on the GLOBALCONFIG SYSPLEXMONITOR parameter.

When should IPsec sysplex monitoring be enabled?

Consider enabling IPsec sysplex monitoring for a TCP/IP stack that has IPsec-protected TCP workloads that are distributed or backed up within the sysplex. A TCP/IP stack that serves only as a target for distributed workloads does not benefit from the monitoring function.

IPsec sysplex monitoring should only be enabled for a stack that has:

- IP security enabled with the IPSECURITY parameter on the IPCONFIG statement in the TCP/IP profile
- Sysplex-wide security associations (SWSA) enabled with the DVIPSEC parameter on the IPSEC statement in the TCP/IP profile
- IKED started on the system

IPsec sysplex monitoring should NOT be used in the following cases:

- IPsec is only used to protect Enterprise Extender (EE) traffic for this stack.

  The actions that the Sysplex Monitor offers rely on DVIPA addresses which are not typically used with EE.
- IKED to NSSD connection uses DVIPA addresses.

  If NSSD is being used for IPsec certificate services and the IKED to NSSD connection uses a DVIPA as the source or destination IP address, then IPsec sysplex monitoring should not be used. IPsec sysplex monitoring requires that IKED successfully connect to NSSD before the stack joins the sysplex group, but DVIPA addresses are not available until the stack has joined the sysplex group. This could lead to a case where the stack is not able to join the sysplex.

  IKED uses the stack's default address as the source address for the TCP connection to NSSD. The destination address is configured on the NssSecurityServer or NssSecurityServerBackup statement in the IKE configuration file.
- Policy Client to Policy Server connection uses DVIPA addresses.

  If a centralized policy server is being used and the policy client and policy server connection uses a DVIPA as the source or destination address, then IPsec sysplex monitoring should not be used. IPsec sysplex monitoring requires that the policy client successfully connect to the policy server and load the policy, but DVIPA addresses are not available until the stack has joined the sysplex group. This could lead to a case where the stack is not able to join the sysplex.

  The policy client uses the stack's default address as the source address for the TCP connection to the policy server. The destination address is configured on the ServerHost or ServerHostBackup parameter on the ServerConnection statement in the policy agent configuration file.
- A limited amount of sysplex traffic is IPsec protected.

  It might be useful to enable DELAYJOINIPSEC so that the stack does not join the sysplex until the IPsec infrastructure is active. However, you might choose NOMONIPSEC to prevent the stack from leaving the sysplex for an IPsec infrastructure failure that only impacts a limited number of connections.

When DELAYJOINIPSEC is configured for a stack, AUTOLOG DELAYSTART DVIPA should not be configured for your Policy Agent, IKED, or NSSD procedures. Note that Policy Agent, IKED, and NSSD are generic servers (those without affinity to a specific stack) and AUTOLOG would not typically be used to start them.

What is monitored in the IPsec infrastructure?

The IPsec infrastructure monitored by sysplex problem detection and recovery (SPDR) includes the Policy Agent, the Internet Key Exchange (IKE) daemon, and the Network Security Server (NSS) daemon (if configured). The following checking determines if the IPsec infrastructure is considered active:

1. IP security is enabled with the IPSECURITY parameter on the IPCONFIG statement in the TCP/IP profile.
2. Sysplex-wide security associations (SWSA) is enabled with the DVIPSEC parameter on the IPSEC statement in the TCP/IP profile.
3. IPsec policy is installed and active for the TCP/IP stack.

4. IKED is active and was able to establish UDP sockets for ports 500 and 4500 for use in negotiating tunnels with peers.
5. IPsec policy is installed in the IKE daemon. The policy must include key exchange policy and dynamic VPN actions which are used for tunnel negotiation. While monitoring cannot determine if your IPsec policy is accurate for individual tunnel negotiations, it ensures that the installed policy contains the appropriate types of policy used for tunnel negotiations.
6. If the IKE daemon is providing certificate services, the IKE key ring can be accessed. Again, monitoring cannot determine if the certificates on your key ring are correct for the tunnels you plan to negotiate. But it ensures that the key ring itself can be accessed and that there is at least one end entity certificate available on the key ring.
7. If the Network Security Server (NSS) is providing certificate services:
   - IKED is connected to NSS and is protected by a TLS session.
   - IKED successfully authenticated using the userid and password/passticket configured in the IKE configuration file.
   - Certificate services are enabled in the IKED configuration file, and permission is defined for the userid to read the EZB.NSS.*sysname.clientname*.IPSEC.CERT SERVAUTH profile.
   - The NSS key ring can be accessed. Again, monitoring cannot determine if the certificates on your key ring are correct for the tunnels you plan to negotiate. But it ensures that the key ring itself can be accessed and that NSS has provided at least one certificate authority certificate for use in negotiations.

Tip: If you are only using pre-shared keys to authenticate all IKE peers then your IPsec infrastructure does not need a keyring or certificates. You can add the NoKeyRing parameter to your IKE configuration file to indicate that no keyring/certificate checking is needed. Checks 6 and 7 are skipped.

Restriction: If both a primary and backup NSSD are configured in the IKED configuration file (NetworkSecurityServer and NetworkSecurityServerBackup), IPsec sysplex monitoring treats NSSD as a centralized resource. The following is done:

- For DELAYJOINIPSEC processing, check 7 above is still applicable.
- For MONIPSEC processing, check 7 is not done. It is likely that a problem with the IKED to NSSD connection for a centralized NSSD affects multiple stacks in the sysplex. Taking a recovery action, such as leaving the sysplex, is not likely to allow a backup stack to re-establish the IPsec tunnels.

What messages are generated when monitoring for IPsec infrastructure issues?

When DELAYJOINIPSEC is configured but sysplex-wide security associations (SWSA) is not enabled for the TCP/IP stack, message EZZ0839I THE DELAYJOINIPSEC PARAMETER ON THE GLOBALCONFIG SYSPLEXMONITOR STATEMENT IS IGNORED BECAUSE SWSA IS NOT ENABLED is generated.

Note: SWSA is enabled with the DVIPSEC parameter on the IPSEC statement block in the TCP/IP profile.

When DELAYJOINIPSEC is configured and the TCP/IP stack is waiting for the IPsec infrastructure to become active, you see eventual action message EZD1976E *stackname* DELAYING SYSPLEX PROFILE PROCESSING – IPSEC INFRASTRUCTURE IS NOT ACTIVE.

When MONIPSEC is configured and an IPsec infrastructure problem is detected after the TCP/IP stack has joined the sysplex, you see eventual action message EZD1977E *stackname* HAS DETERMINED THAT THE IPSEC INFRASTRUCTURE WAS NOT RESPONSIVE FOR AT LEAST *timersecs* SECONDS.

When MONIPSEC is configured, if the IPsec default filter rules are activated, you see eventual action message EZD1979E *stackname* HAS DETERMINED THAT THE DEFAULT IPSEC FILTERS HAVE BEEN IN EFFECT FOR AT LEAST *timersecs* SECONDS.

How do I know why I am seeing message EZD1976E or EZD1977E?

When EZD1976E or EZD1977E is present, syslogd should be reviewed for IKED message EZD2050I IPSec infrastructure problem detected for stack *stackname : reason*. The *reason* value provides a more detailed explanation of the detected problem.

The TCP/IP stack relies on IKED to provide a heartbeat indicating that the IKED and NSSD infrastructure is active (checks 4-7 above). IKED withholds the heartbeat if a problem is detected and generates an EZD2050I message. As the TCP/IP stack, Policy Agent, IKED, and NSSD are starting, IKED might generate an EZD2050I message for a condition that is cleared once all resources are active. IKED generates an EZD2049I message when a problem condition is cleared. For more information on messages EZD2049I and EZD2050I, see z/OS Communications Server: IP Messages Volume 2 (EZB, EZD).

Depending on the *reason* reported in EZD2050I, syslogd should be reviewed to determine what errors are being reported. Errors could be specific to IKED or NSSD. Errors could be related to establishing the AT-TLS connection required between IKED and NSSD, or errors could be related to resource permissions.

Target server connection setup responsiveness monitoring

Sysplex distributor measures the responsiveness of target servers in accepting new TCP connection requests at intervals of approximately 1 minute. Target server responsiveness (TSR) values calculated from these measurements are used to modify the weight when using WLM system weight distribution, WLM server-specific weight distribution, or weighted active distribution, to favor those servers that are more successfully accepting new TCP connection setup requests.

TSR

TSR values are displayed as a percentage. Thus, a value of 100 indicates full responsiveness and a value of 0 indicates no responsiveness in setting up new TCP connections. A value of 0 for the TSR causes this target server to be bypassed when new TCP connection setup requests are distributed to target servers for a DVIPA and port, even if the distribution method is round-robin. Periodically, the distributor sends a new connection request to a target with a TSR of 0, to check whether the responsiveness of that target has improved.

The calculated TSR values are applied to the WLM weights only if there are no stacks participating in the distribution of connections for this DVIPA that are at a release level earlier than z/OS V1R7. The TSRs are calculated by combining two factors, the target connectivity success rate (TCSR) and the Servers' accept Efficiency Fraction (SEF). These factors are defined as follows:

TCSR
The target connectivity success rate is the measure of how many new TCP connection setup requests (SYNs) sent by the distributing stack to a target stack are received by the target stack. A low value could indicate a problem with the connectivity between the distributing stack and the target stack.

SEF
The Servers' accept Efficiency Fraction measures how well a target server is accepting new TCP connection setup requests and managing its backlog queue.

CER

The connection establishment rate (CER) measures how many TCP connection setup requests received by the target stack become established (that is, how many enter the ESTAB state). The CER value is displayed as a percentage; 100 indicates that all of the connection setup requests that were received entered the established state. A lower value could indicate a routing problem in the network. The CER value is shown only as a diagnostic aid, and its value has no affect on the WLM weight.

Workload balancing

Load balancing is the ability for a cluster to distribute workload evenly (or based on some policy) to target servers comprising the cluster. Usually, this load balancing is calculated based on the perceived load on

each of the target servers. Various techniques are available to perform IP load balancing for a System z sysplex. These techniques typically fall into the following categories:

- Internal load balancing solutions

  These solutions typically rely on a component executing within the z/OS sysplex to perform the load balancing with very few dependencies on the external network. Because they are executing within the sysplex environment, these solutions typically have access to information that can be used to optimize load balancing, such as system capacity and current load information. An example of this technique is the sysplex distributor.

- Sysplex-aware external load balancing solutions

  These solutions are typically comprised of an external load balancer that relies on components inside the z/OS sysplex for advice on how workload should be distributed. As a result, these solutions usually provide for load balancing that is optimized for a z/OS sysplex environment. Examples of these solutions include the IBM Network Dispatcher and any external load balancing solution that supports the Server/Application State Protocol (SASP) provided by the z/OS Load Balancing Advisor (for example, the CISCO Content Switching Module).

- External load balancing solutions

  These solutions are usually comprised of load balancing components that execute outside of the z/OS sysplex, typically on one or more hosts or routers in the network, with little or no specific awareness of the z/OS sysplex environment. Several vendors provide such load balancing solutions.

This information describes and compares these three load balancing techniques and selected associated solutions. Each solution identifies the target System z servers that can receive client connections based on some specification.

By providing load balancing, clustering techniques must also provide for other system requirements in addition to the dispatching of connections. These include the ability to advertise some single systemwide image or identity so that clients can uniquely and easily identify the service. Additionally, clustering techniques should also provide for horizontal growth of the system and ease of management.

Single systemwide image

Clients connecting to a cluster should not be aware of the internal makeup of a cluster. More specifically, clients should not even be aware that the service they are requesting is actually being serviced by a collection or cluster of servers. Instead, clients must be provided with some single image identifier to be used when connecting to the service. For example, most of the load balancing solutions (internal or external) discussed in this information use a single IP address to represent a cluster of servers to clients. The clients use this IP address to establish connections and are not aware that the load balancing solution directs requests to a specific instance of the server.

Horizontal growth

As the clients' demands on the service increase, clusters must provide a way to expand the cluster of servers to accommodate for such growing demand. Put in another way, the cluster must provide a mechanism by which to add servers without disrupting the operation of the cluster. To this end, the service is made available to clients at all times and can grow horizontally to accommodate for increased demand placed on the cluster by the clients.

Ease of management

The administrative burden associated with the cluster (sysplex) should not increase as you add servers to the cluster. You should use the same configurations for many systems in the cluster. Within a sysplex, servers are homogenous, because a sysplex is comprised solely of System z servers. As such, many of the configurations can be shared among the different System z servers, which reduces the administrative burden associated with the sysplex. Additionally, as the size of the cluster increases, the administrative overhead in adding systems to the cluster should be as low as possible.

Administrative activities required to maintain a load balancing solution are highly dependent on the type of load balancing solution deployed. For example, administrative tasks associated with the maintenance of internal load balancing solutions typically consist of administrative tasks within the z/OS sysplex, such as changing z/OS configuration files. External load balancing solutions typically require administrative tasks performed on the external load balancing components, while sysplex-aware external load balancing solutions might require administrative tasks on both internal and external load balancing components. Depending on your environment and organizational structure, the administrative scope required by the various load balancing solutions can play a key role in your selection process. For example, do you need to deploy a solution that requires only administrative tasks on z/OS, or can you make required updates to network components as necessary?

Internal load balancing solutions

Internal load balancing solutions typically rely on components that are within the z/OS sysplex to perform the load balancing. While other components might be required in the network, after the solution is configured, minimal changes should be needed from network components outside of the sysplex. An example of an internal load balancing solution is the sysplex distributor. For information about the sysplex distributor, see "Sysplex distributor" on page 475.

Sysplex-aware external load balancing solutions

Sysplex-aware external load balancing solutions are typically comprised of an external load balancing solution that communicates with components within the z/OS sysplex, to optimize load balancing based on the current conditions and workloads in the sysplex environment. Examples of these solutions include:

- External load balancers that use the Server/Application State Protocol (SASP) to obtain recommendations and topology information related to applications and systems within a sysplex environment.
- The IBM Network Dispatcher can obtain WLM recommendations from z/OS, and uses these recommendations to determine how workload requests are routed.

External load balancers can obtain detailed information regarding the state of target z/OS applications and systems by communicating with the z/OS Load Balancing Advisor using the SASP protocol. Using SASP, external load balancers can also obtain detailed recommendations on how workload should be distributed within the sysplex environment.

A key component of these recommendations is derived from the z/OS Workload Manager (WLM). The WLM information reflects the current available system capacity for target systems, and can also reflect how well individual server applications are performing compared to the WLM policy goals specified for that application. These recommendations are also influenced by how well the applications are performing from a TCP/IP perspective. For example, are the applications processing new requests quickly enough? Or are new requests filling up the backlog queues that TCP/IP maintains for each application?

Using SASP, external load balancers can perform load balancing that is optimized for your sysplex environment, based on the current configuration and workload conditions. SASP supports TLS/SSL for secure authentication, access control, and encryption of data. For more information on SASP and the z/OS Load Balancing Advisor, see Chapter 21, "z/OS Load Balancing Advisor," on page 1147.

External IP workload balancing solutions

IBM's Network Dispatcher (part of WebSphere Edge Server) and Cisco's Content Switching Module (CSM) are examples of external IP workload balancing solutions. Such solutions exist outside the sysplex, but can direct work into the sysplex. External IP workload balancing solutions typically define a single IP address representing all instances of the server, and then balance new work requests (for example, new TCP connection requests) among available servers.

External load balancers typically use a cluster IP address to represent the set of applications being load balanced. Client applications use this cluster IP address as the destination IP address for all their requests. External load balancers might be capable of using different methods for forwarding packets to their destinations. Various load balancing solutions might use different terms when describing these

methods. In this discussion, dispatch mode and directed mode are discussed as two examples of these methods.

When a load balancer uses dispatch mode, the destination IP addresses for incoming IP packets is not changed. Instead, the load balancer forwards the packet to a target z/OS system by using the MAC address of a network adapter on that system. This technique works well when the target systems and the load balancer are attached to the same network (that is, there are no intervening routers). The receiving z/OS system inspects the destination IP address of the packet, and if it matches one of the IP addresses in its HOME list, it accepts the packet. As a result, with dispatch mode, all target z/OS systems must have the load balancer's cluster IP address defined in their HOME list. It is important, however, that these addresses are not advertised externally through dynamic routing protocols. One way to accomplish this is by defining these IP addresses as loopback addresses on z/OS.

Dispatch mode also has special considerations when the load balancer is more than one hop away from the target systems (that is, a packet must be routed to the target), or when multiple z/OS target systems share the same OSA-Express adapter. In these scenarios, you can use the following techniques to route packets to the correct target system:

- VMAC addresses

  Causes the OSA to route packets using a MAC address that is exclusively owned by the target system.
- GRE tunnels

  Causes the OSA to route packets using an IP address that is exclusively owned by the target system.

On the other hand, when a load balancer uses directed mode, the load balancer converts the destination IP address (that is, the cluster IP address) to an IP address owned by the target z/OS system, using technologies such as network address translation (NAT). When IP packets for these connections are sent back to clients, the load balancer converts the source IP address (that is, the target z/OS system's IP address) back to the cluster IP address that the application had used on its request.

Some load balancer configurations might also perform NAT on the client's IP address. The client's IP address can be replaced with an IP address owned by the load balancing device. This might be necessary to ensure that all outgoing packets from a target system traverse the load balancing device, so that NAT can be performed to change the server's IP address back to the cluster IP address that the client had originally used. Therefore, it is important to note that with directed mode solutions, the IP addresses of load balanced connections reaching the sysplex might not reflect the real IP addresses of the clients making the requests. This can be an important consideration if any definitions or configuration within the sysplex rely on the client's IP address being visible on incoming connections.

External load balancing solutions might provide several configuration options that influence how workload is distributed, such as round-robin, weighted round-robin, and so on. Some of the solutions might also obtain recommendations from components inside the z/OS sysplex that might affect their workload balancing decisions. For more information, see “Sysplex-aware external load balancing solutions” on page 470.

Choosing a load balancing solution

Several load balancing solutions are available for a z/OS environment. The solution that is best for your environment depends on several factors that are related to your environment and the workload being load balanced, including:

- What type of workload needs to be load balanced, and does a particular load balancing solution support this type of workload? For example, does the workload have affinities to specific servers? Can these affinities be determined only by inspecting data content? If so, an external load balancing solution that supports content inspection and affinities might be the most appropriate solution.
- Is it important that the solution is primarily configured and maintained within the z/OS sysplex, with minimal interactions with network components? In these scenarios, an internal load balancing solution might be the preferred answer.

- Do you already have an external load balancing solution in the network? If so, can it be used to load balance z/OS workloads? Does it support SASP, so that it can perform more optimal load balancing decisions?
- Does the load balancing solution have the needed level of availability? For example, if the primary load balancer fails, can a secondary load balancer take over transparently? Can it do so without disrupting existing connections? What if target applications or systems fail? Does the load balancer recognize these conditions?

While a single load balancing solution might be desirable for all workloads in some environments, you can select multiple solutions based on your specific requirements. For example, you might select an external load balancing solution for a particular workload, while an internal load balancing solution might be enabled for another workload. Table 26 on page 472 lists some of the key attributes of the various load balancing techniques and specific solutions that were discussed in this information. It can be used as a quick reference for comparing these solutions.

Table 26. Load balancing solution quick reference

Features or considerations	**Sysplex distributor**	**External load balancers with SASP**	**External load balancers**
How is the solution configured and administered?	Initial setup might require some interaction with the network (dynamic routing protocols, DNS updates for dynamic VIPAs, and so on). Ongoing administration (adding and removing target server applications and systems) typically confined within z/OS systems.	Initial setup and configuration on load balancer and on z/OS. Ongoing administration might need to be performed on both the load balancer and the z/OS systems.	Initial setup and configuration on load balancer, some configuration on z/OS might be required. Ongoing administration should be mostly confined to the load balancer, although z/OS configuration might be necessary, such as when adding new target systems.
When is the server instance decision made?	Connection setup (in line Syn segment)	Connection setup (in line Syn segment)	Connection setup (in line Syn segment)
Support for TCP and UDP applications?	TCP only	Depends on the load balancer implementation; SASP supports both TCP and UDP	Depends on the load balancer implementation
Extra network flows?	No, for outbound traffic. Yes, for inbound traffic. Inbound traffic must traverse the sysplex distributor node. If sysplex distributor is configured as service manager for CISCO routers, the inbound traffic can flow directly to the target application.	Depends on the load balancer implementation; can be avoided if the load balancer is implemented as part of a router or switch.	Depends on the load balancer implementation; can be avoided if the load balancer is implemented as part of a router or switch.

Table 26. Load balancing solution quick reference (continued)

Features or considerations	Sysplex distributor	External load balancers with SASP	External load balancers
Support for affinities between TCP connection requests based on data content?	No, but support does exist for timer-based affinities	Depends on implementation; some support affinities for HTTP and HTTPS requests by inspecting data content (correlating cookies, jsessionid)	Depends on implementation; some support affinities for HTTP and HTTPS requests by inspecting data content (correlating cookies, jsessionid)
Network address translation?	Not needed; client and server IP addresses are not modified	Might be required by some implementations; client and server IP addresses might be translated	Might be required by some implementations; client and server IP addresses might be translated
Support for IPv6?	Yes	Depends on the load balancer implementation; SASP supports both IPv4 and IPv6	Depends on the load balancer implementation
z/OS WLM recommendations?	Yes, system level and server level	Yes, system level and server level	Depends on the load balancer implementation
z/OS network QoS recommendations?	Yes, based on z/OS QoS policy	No	No
z/OS TCP/IP server health information?	Yes	Yes	No

Table 26. Load balancing solution quick reference (continued)

Features or considerations	Sysplex distributor	External load balancers with SASP	External load balancers
Detection of target application and target system state changes, active or inactive?	Yes, application and system state changes are detected in near real-time fashion.	Yes, the z/OS Load Balancing Advisor and Agents detect application and system state changes within a configurable time period, 60 seconds by default. How quickly external load balancers become aware of these changes depends on several factors: • If the load balancer is using a push model with SASP, the Load Balancing Advisor sends a notification of a state change as soon as it is detected. • If the load balancer is using a poll model with SASP, it depends on the load balancer's polling interval. • The load balancer might also have additional mechanisms for detecting application and system state changes, which might provide for faster detection of these changes.	Depends on the load balancer implementation.
High availability solution, load balancing continues even if the primary load balancing component becomes unavailable?	Yes, one or more backups can be configured to enable dynamic takeover in cases where the TCP/IP stack, or system that is acting as the distributor, fails.	For failures to the load balancer, it depends on the load balancer implementation. Some solutions provide for backup load balancers that can dynamically take over load balancing responsibilities in cases of failures. The z/OS Load Balancing Advisor and Agents can be configured for high availability to minimize the impact of an Advisor, Agent, or system failure.	For failures to the load balancer, it depends on the load balancer implementation. Some solutions provide for backup load balancers that can dynamically take over load balancing responsibilities in cases of failures.

Table 26. Load balancing solution quick reference (continued)			
Features or considerations	**Sysplex distributor**	**External load balancers with SASP**	**External load balancers**
Caching issues?	No, every new connection request is eligible for load balancing.	No, every new connection request is eligible for load balancing.	No, every new connection request is eligible for load balancing.

Sysplex distributor

Sysplex distributor extends the notion of dynamic VIPA and automatic VIPA takeover to allow for load distribution among target servers within the sysplex. It extends the capabilities of dynamic VIPAs to enable distribution of incoming TCP connections to ensure high availability of a particular service within the sysplex.

The functionality of sysplex distributor is that one IP entity advertises ownership of an IP address by which a particular service is known. In this fashion, the single system image of sysplex distributor is that of a special IP address. This IP address is called a distributed DVIPA. Further, in sysplex distributor, the IP entity advertising the distributed DVIPA and dispatching connections destined for it is itself a system image within the sysplex, referred to as the distributing stack.

Sysplex distributor makes use of Workload Manager (WLM) and its ability to gauge server load and provide a WLM recommendation. In this paradigm, WLM provides the distributing stack with a WLM recommendation for each target system (a WLM system weight) , or the target stacks provide the distributing stack with a WLM recommendation for each target server (a WLM server-specific weight). The distributing stack uses this information to optimally distribute incoming connection requests between a set of available servers. Additionally, sysplex distributor has the ability to specify certain policies within the Policy Agent so that it can use QoS information from target stacks to further modify the WLM recommendation. Further, these policies can specify which target stacks are candidates for clients in particular subnetworks.

Sysplex distributor also measures the responsiveness of target servers in accepting new TCP connection setup requests, favoring those servers that are more successfully accepting new requests. For more information, see “Target server connection setup responsiveness monitoring” on page 468.

The Workload Manager weight is a relative value. The WLM system weight is an indication of the target system's capacity for additional work, and the WLM server-specific weight is a more granular indication of the specific target server's capacity for additional work. Higher numbers indicate a target with comparatively greater capacity. The Netstat VDPT/-O display shows these values. WLM system weights are indicated by a B beside the weight, and WLM server-specific weights are indicated by an S beside the weight.

Distribution using WLM system weights is the default distribution method. This can be specified by using the BASEWLM parameter on the VIPADISTRIBUTE statement. Distribution using WLM server-specific weights can be specified by using the SERVERWLM parameter on the VIPADISTRIBUTE statement. If the SERVERWLM parameter is used and all stacks are able to provide WLM server-specific weights for that VIPA/port, WLM server-specific weights are used. Otherwise, WLM system weights are used.

BASEWLM - Distribution using WLM system weights

When determining a WLM system weight recommendation, WLM assigns a relative weight to each system in the sysplex, with the highest weight going to the system with the most available capacity. The available capacity is based on the system's available general CPU capacity, and optionally (if there are no systems in the sysplex that are prior to V1R9), available System z Application Assist Processor (zAAP) capacity and available System z Integrated Information Processor (zIIP) capacity. When PROCTYPE for BASEWLM is configured on the VIPADISTRIBUTE statement, a composite WLM weight is determined by the sysplex distributor based on the WLM weight for each processor type and expected usage of each processor type by an application.

If all systems in the sysplex are running at or near 100%, WLM assigns the highest weights to the systems with the largest amount of lower importance work. In this way, new connection requests are distributed to the systems with the highest *displaceable* capacity. However, this assumes that the new work is of a high enough importance level to displace this work. A system can be so loaded that only higher importance work is running on that system, in which case the new work request is not able to meet the goals specified in the WLM policy for that server. For more information on how displaceable capacity is calculated and using the Sysplex Routing Services, see z/OS MVS Programming: Workload Management Services.

Use the Netstat VDPT/-O report option with the DETAIL modifier to determine the WLM system weight recommendations for each processor type, along with the modified weight for each processor based on the expected usage proportion configured on the VIPADISTRIBUTE statement.

SERVERWLM - Distribution using WLM server-specific weights

When WLM determines a WLM server-specific recommendation, it determines the service class of the server address space and then assigns a weight that is based on the following criteria:

- How well that server is actually meeting the goals of its service class.
- The amount of displaceable workload available, given the importance of the service class.
- If the distributor and target systems are not prior to V1R9, the WLM recommendation is a composite weight based on the available and displaceable capacity of each processor type (general, zAAP, and zIIP) and the current usage of each processor type by this application.

If the distributor and target systems are not running a release prior to V1R11, you can use the PROCXCOST and ILWEIGHTING parameters on the VIPADISTRIBUTE statement to influence the WLM server-specific recommendation.

- PROCXCOST

  For applications that are designed to have a portion of their workload run on a zAAP or zIIP specialty processor, use the PROCXCOST parameter on the VIPADISTRIBUTE statement to influence how aggressively WLM favors servers on systems with available zAAP or zIIP capacity over servers on systems where work targeted for the specialty processors has instead run on the conventional processor. The PROCXCOST parameter specifies a crossover cost in the range 1–100 that is applied to the amount of zAAP-targeted or zIIP-targeted workload that actually ran on the conventional processor. The resulting cost is used to reduce conventional processor use by the application, which in turn reduces the WLM recommendation for that server. A crossover cost of 1 means that zAAP-targeted or zIIP-targeted workload that runs on the conventional processor should not be penalized; the WLM recommendation is not reduced.

  Before specifying a high crossover cost, give careful consideration to the effect this might have on workload distribution, because servers with less zAAP or zIIP processor capacity might receive a significantly lower percentage of the overall workload. The RMF Workload Activity Report shows the zAAP and zIIP processor use, as well as how much crossover took place. Run this report before and after using the PROCXCOST parameter to better understand how it affects your overall workload performance.
- ILWEIGHTING

  Use the ILWEIGHTING parameter on the VIPADISTRIBUTE statement to influence how aggressively WLM favors servers on systems with displaceable capacity at lower importance levels over servers on systems with displaceable capacity at higher importance levels. Work is categorized into one of seven importance levels (importance levels 0 through 6). The most important work runs at importance level 0 and the least important work runs at importance level 6.

  The ILWEIGHTING parameter value specifies how WLM should consider displaceable capacity when determining a server-specific recommendation. The higher the value specified for ILWEIGHTING, the more a stack with displaceable capacity at lower importance levels is favored. The possible ILWEIGHTING parameter values and examples of their effect are as follows:

  – ILWEIGHTING 0

Specifies that WLM should ignore importance levels when comparing displaceable capacity. This is the default value. For example, if system A and system B are running at 100 percent capacity and new work will be running at importance level 2, all work at importance level 3 and greater is considered to be available displaceable capacity.

- If system A has 500 service units of work at importance level 3 that can be displaced as the new work runs on this system, it initially has a raw WLM weight of 500.
- If system B has 500 service units of work at importance level 6 that can be displaced, it also has an initial raw WLM weight of 500.

The WLM weights are initially the same for both systems, but might be different after the health and performance of the application are considered. The actual WLM weights returned to the invoker are normalized by WLM so that they are in the range 0–64.

In this example, as 100 connections are received, 50 connections are distributed to system A and 50 connections are distributed to system B, assuming that the health and performance of each application are optimal.

– ILWEIGHTING 1

Specifies that WLM should weigh displaceable capacity at each successively lower importance level slightly higher than the capacity at the preceding higher importance level. The weighting grows proportionally to the square root of the importance level difference plus 1. This provides a moderate bias when comparing displaceable capacity at different importance levels. Using the previous example but with ILWEIGHTING 1 specified, system B is preferred over system A even though 500 service units of work are to be displaced on either system.

- The raw WLM weight of system A is 707:

```
  500 × SQUAREROOT ((Importance level of displaced work - Importance level of new work) +
1)
= 500 × SQUAREROOT ((3 - 2) + 1)
```

- The raw WLM weight of system B is 1118:

```
  500 × SQUAREROOT ((Importance level of displaced work - Importance level of new work) +
1)
= 500 × SQUAREROOT ((6 - 2) + 1)
```

In this example, as 100 connections are received, roughly 39 are distributed to system A and 61 are distributed to system B, assuming that the health and performance of each application are optimal.

Guideline: If you are using an ILWEIGHTING value for the first time other than the default value 0, use the value 1 initially.

– ILWEIGHTING 2

Specifies that WLM should weigh displaceable capacity at each successively lower importance level significantly higher than the capacity at the preceding higher importance level. The weighting grows proportionally to the importance level difference plus 1. This provides an aggressive bias when comparing displaceable capacity at different importance levels. Using the original example but with ILWEIGHTING 2 specified:

- The raw WLM weight of system A is 1000:

```
  500 × ((Importance level of displaced work - Importance level of new work) +
1)
= 500 × ((3 - 2) + 1)
```

- The raw WLM weight of system B is 2500:

```
  500 × ((Importance level of displaced work - Importance level of new work) +
1)
= 500 × ((6 - 2) + 1)
```

In this example, as 100 connections are received, roughly 28 are distributed to system A and 72 are distributed to system B, assuming that the health and performance of each application are optimal; more than twice as many are routed to system B than system A.

– ILWEIGHTING 3

Specifies that WLM should weigh displaceable capacity at each successively lower importance level significantly higher than the capacity at the preceding higher importance level. The weighting grows proportionally to the square of the importance level difference plus 1. This provides an exceptionally aggressive bias when comparing displaceable capacity at different importance levels. Using our original example but with ILWEIGHTING 3 specified:

- The raw WLM weight of system A is 2000:

```
  500 × SQUARE ((Importance level of displaced work - Importance level of new
work) + 1)
= 500 × SQUARE ((3 - 2) + 1)
```

- The raw WLM weight of system B is 12500:

```
  500 × SQUARE ((Importance level of displaced work - Importance level of new
work) + 1)
= 500 × SQUARE ((6 - 2) + 1)
```

In this example, as 100 connections are received, roughly 14 are distributed to system A and 86 are distributed to system B, assuming that the health and performance of each application are optimal; more than six times as many are routed to system B than system A.

Use the Netstat VDPT/-O report option with the DETAIL modifier to determine the WLM server-specific weight recommendations for each processor type, along with the modified weight for each processor based on the current usage of each processor type.

Evaluate whether WLM server-specific weight distribution can be used as an alternative to WLM system weight distribution for an application. In addition to the reasons mentioned previously, WLM server-specific weight distribution has the added advantage that processor proportions are automatically determined and dynamically updated by WLM based on the actual CPU usage by the application. If you need to determine the processor proportions necessary for WLM system weight distribution, study the workload usage of assist processors by analyzing SMF records, use performance monitor reports such as RMF, and so on.

While WLM server-specific recommendations can be very effective in helping load balancers optimize their routing decisions, there are some scenarios in which applications that are load balancing targets might experience issues that WLM is not aware of through its normal monitoring functions. For example, consider a scenario in which a specific server application instance is executing on a system with excess CPU capacity, yet it cannot successfully process any transactions routed to it because the back-end database it requires on that system is not currently available. From a WLM perspective, this server application instance appears to be a very good candidate to receive requests, given the current available capacity of the system, and the fact that transactions routed to this server appear to be completing very quickly and consuming very little CPU capacity. WLM would therefore assign a higher server-specific weight to this server instance, causing more work to be routed to the ailing server. This type of problem is sometimes referred to as a storm drain problem.

To help alleviate this problem, WLM also considers the health of the server application in its server-specific recommendations. The health of the server is directly determined by information that the server application provides to WLM through programming interfaces. In this way, the WLM perceived health of the application depends directly on the amount of information that an application provides to WLM, if any. WLM uses this information to potentially reduce the weight recommendations for specific server applications that are experiencing problems. This enables sysplex distributor to direct fewer new TCP connections to these servers, selecting instead servers that are not experiencing these problems. WLM considers the following components when determining a server application's health:

- Rate of abnormal transaction completions reported by the server application

This is applicable to applications, such as the CICS Transaction Server for z/OS, that act as Subsystem Work Managers, reporting transaction status using Workload Management Services, such as IWMRPT. For example, if an application reports to WLM that 90% of all transactions it processed completed abnormally, this server is probably not a good candidate for receiving many new TCP connection requests, as these requests will likely also complete abnormally. As a result, WLM significantly reduces the weight recommendation that is returned to sysplex distributor for this server.

- General health of the application as reported by the server application

  This health indicator is available only for applications that provide this information to WLM using the IWM4HLTH or IWMSRSRG services. The health indicator provides a general health indication for an application or subsystem. Under normal circumstances, the value of this field is 100, meaning the server is 100% healthy. Any value less than 100 indicates that the server is experiencing problem conditions that are not allowing it to process new work requests successfully. A value of less than 100 also causes the WLM to reduce the recommendation provided to sysplex distributor for this server instance.

Choosing between the BASEWLM and SERVERWLM distribution methods

For most applications, the WLM server-specific recommendations provide a more accurate way to distribute workload to the servers. However, when a server acts as an access point to applications that run in other address spaces (and therefore in a different service class), WLM system weights might be the preferred distribution method. For example:

- The TN3270E Telnet server (Telnet) is a communication gateway function that enables clients to access SNA applications over an IP network. As a result, most of the actual work associated with a Telnet workload takes place in the SNA application, which is typically classified to a different WLM service class than Telnet and probably at a lower importance level. Therefore, although the WLM server-specific recommendation can provide an accurate assessment of how well an individual TN3270E Telnet server is performing, it does not necessarily provide an accurate assessment of the available capacity required by the back-end SNA applications that the client is accessing. As a result, WLM system weight is probably a more appropriate distribution method for this server.
- The INET daemon also provides access to other applications that are probably associated with service classes of a lower importance level (for example, z/OS UNIX Telnet, REXECD, and RSHD). As a result, similar considerations apply, and WLM system weights are the more appropriate distribution method for this server.
- The FTP daemon address space typically performs very little processing on behalf of new FTP sessions. After accepting a new connection, it performs a fork() for an FTP server process that will service the new FTP session. The FTP server process is classified again by WLM, possibly resulting in a different service class than that of the FTP daemon.
  - If the FTP servers run in a different service class than the FTP daemon, then WLM system weights should be used.
  - If WLM policies are set up so that the FTP servers are classified in the same service class as the FTP daemon, WLM server-specific weights should be the distribution method. Although the WLM recommendation will not take into account how well the actual FTP server is meeting its goal, the recommendation will more accurately reflect the amount of displaceable capacity because it is based on the importance of the service class.

BASEWLM and SERVERWLM display example

In the following Netstat VDPT/-O example, the weights represent normalized weights. That is, the original raw weights received from WLM are proportionally reduced for use by the distribution algorithm. Connections are distributed to these servers in a weighted, round-robin fashion using the normalized weights. In this example, the target server responsiveness (TSR) values are all 100, indicating that all servers are fully responsive to new connection requests. A value of 100 is also displayed for stacks that are at a level prior to z/OS V1R7. In that case, no target server responsiveness calculations are applied to the WLM values.

The SERVERWLM parameter was not coded for DVIPA 201.2.10.11, port 245, so it is using WLM system weights.

```
MVS TCP/IP NETSTAT CS VxRx       TCPIP NAME: TCPCS           12:19:18
Dynamic VIPA Distribution Port Table:
Dest IPaddr     DPort DestXCF Addr    Rdy TotalConn  WLM  TSR Flg
-----------     ----- ------------    --- ---------  ---  --- ---
201.2.10.11     00245 201.1.10.10     001 0000000000 12   100 DB
201.2.10.11     00245 201.1.10.15     001 0000000000 04   100 B
201.2.10.12     04011 201.1.10.10     001 0000000000 04   100 S
201.2.10.12     04011 201.1.10.15     001 0000000000 08   100 S
201.2.10.12     04011 201.1.10.40     001 0000000000 16   100 S
```

If the BASEWLM parameter or the SERVERWLM parameter is specified and WLM weights are not available, incoming connections are distributed among all available servers using round-robin distribution.

Now consider the same example, as shown in the following Netstat VDPT/-O display, but with TSR values indicating that some of the servers are not accepting new connection setup requests productively. The displayed WLM weights have been modified by the TSR values:

```
MVS TCP/IP NETSTAT CS VxRx       TCPIP NAME: TCPCS           12:19:18
Dynamic VIPA Distribution Port Table:
Dest IPaddr     DPort DestXCF Addr    Rdy TotalConn  WLM  TSR Flg
-----------     ----- ------------    --- ---------  ---  --- ---
201.2.10.11     00245 201.1.10.10     001 0000000000 09   075 DB
201.2.10.11     00245 201.1.10.15     001 0000000000 04   100 B
201.2.10.12     04011 201.1.10.10     001 0000000000 03   090 S
201.2.10.12     04011 201.1.10.15     001 0000000000 06   075 S
201.2.10.12     04011 201.1.10.40     001 0000000000 03   020 S
```

The SERVERWLM parameter was coded for DVIPA 201.2.10.12, port 4011. For the server on destination 201.1.10.10, a TSR value of 90 indicates that the server is 90% responsive in accepting new connection requests. For the server on 201.1.10.15, a TSR value of 75 indicates that this server is 75% responsive in handling new connection requests. Likewise, for the server on 201.1.10.40, a TSR value of 20 indicates that it is only 20% effective in accepting new connection requests. These factors are used to modify the WLM server-specific weights, and the modified weights are normalized. As a result, the server on destination XCF 201.1.10.10 has a normalized WLM value of 3, the server on XCF 201.1.10.40 has a WLM value of 3, and the server on destination XCF 201.1.10.15 has a WLM value of 6. So, the server on destination XCF 201.1.10.15 is now favored over the other servers, and the server at 201.1.10.40 is now equal to the server at 201.1.10.10. Connections are distributed to these servers in a weighted, round-robin fashion using these normalized weights.

The SERVERWLM parameter was not coded for DVIPA 201.2.10.11 port 245, so it is using WLM system weights. The server at 201.1.10.10 has a TSR of 75, so the normalized weight is 9, three-quarters of what it was in the previous example. The server at 201.1.10.15 has a TSR of 100, so the normalized weight is the same as it was in the previous example.

WEIGHTEDACTIVE - Distribution based on active connection load

In some instances, rather than using WLM recommendations, weighted active connections (WEIGHTEDActive) can provide a more appropriate solution to control workload distribution:

- Application scaling concerns

  Target systems vary significantly in terms of capacity (small systems and larger systems). WLM recommendations might favor the larger systems significantly. However, a target application might not scale well to larger systems; because of its design, it might not be able to take full advantage of the additional CPU capacity on the larger systems. This can result in these types of servers getting inflated WLM recommendations when running on larger systems and getting overloaded with work.

- Unequal number of SHAREPORT servers

  SHAREPORT is being used, but not all systems have the same number of SHAREPORT server instances (for example, one system has two instances and the other has three). The current round-robin or WLM recommendations do not change distribution based on the number of server instances on each target. Round-robin distribution distributes one connection per target stack regardless of the number

of SHAREPORT server instances on that stack. WLM server-specific weights from a target stack with multiple server instances reflect the average weight.

- You need to control the amount of capacity that specific workloads can consume on each system.

  For example, you might need to reserve some capacity on certain systems for batch workloads that are added into selected systems during specific time periods and have specific time window completion requirements. If those systems are also a target for long-running distributed DVIPA connections, WLM recommendations allow that available capacity to be consumed. This can potentially impact the completion times of the batch jobs when they begin to run, if they are not able to displace the existing non-batch workloads. Similarly, the existing connections on that system can experience performance problems if the batch jobs displace those workloads.

Weighted active connections provide granular control over workload distribution based on predetermined active connection count proportions for each target (fixed weights). Distribution of incoming TCP connection requests is balanced across the targets so that the number of active connections on each target is proportionally equivalent to a configured active connection weight for each target (specified on the DESTIP parameter for each target). Control is gained at the expense of losing the dynamic benefits of WLM recommendations; however, server-specific abnormal completion information, the general health indicator, and the TSR value are used to reduce the active connection weight when these indicators are not optimal. If weighted active connections are used, study and determine the comparative workload that you want on each system so that you can configure appropriate connection weights.

To enable the distributing stack to use server-specific abnormal completion and health information to affect the active connection weight, specify SYSPLEXROUTING on the IPCONFIG statement for all participating stacks.

You can select this type of distribution for DVIPA and port targets by specifying the WEIGHTEDActive option on the DISTM parameter of the VIPADISTRIBUTE statement, and specifying the proportion of active connections that you want on each target using the WEIGHT option on the DESTIP *dynxcfip* parameter of the VIPADISTRIBUTE statement.

Choosing between RoundRobin and WeightedActive distribution

If weighted active connections are configured with default proportions for each target, connections are evenly distributed across all target servers; the goal is to have an equal number of active connections on each server. This is similar to the round-robin distribution method; the difference is that round-robin distribution distributes connections evenly to all target servers without consideration for the active number of connections on each server.

In certain scenarios, a round-robin distribution might be appropriate for specific server applications. Select this type of distribution for particular targets by specifying the DISTM ROUNDROBIN parameter on the VIPADISTRIBUTE statement.

Although both round-robin and weighted active connection distribution do not consider WLM recommendations in their target selection, weighted active connection distribution does use the server-specific abnormal completion information, the general health indicator, and the TSR value to reduce the weight when these values are less than optimal. For round-robin distribution, connections are not distributed to a target if its TSR value is 0. Periodically, the distributor sends a new connection request to a target with a TSR of 0 to check whether the responsiveness of that target has improved.

Hot standby distribution

You can use the hot standby distribution method to configure sysplex distributor to have one preferred target server and one or more backup (hot standby) target servers. In this configuration, sysplex distributor does not perform load balancing of new connection requests across multiple targets; rather, a preferred target server with an active listener receives all new incoming connection requests. The hot standby target servers, which typically also have a ready listener application, do not receive any new connections requests; they act as backup target servers in case the designated preferred target server become unavailable. A target is considered unavailable if any of the following conditions are true:

- The target is not ready.

- The distributor does not have an active route to the target.
- The target is not healthy.

The hot standby distribution method is useful in situations where there is a trade-off between availability and performance, such as a scenario where data sharing between multiple server applications and LPARs is required for availability, but it is more efficient if the work runs on one LPAR instead of interlocking access to the same data across multiple LPARs. For example, if you prefer that an application instance on one LPAR, if available, always handle the workload, and an application instance on another LPAR that is otherwise performing lower priority work be available for backup, use the hot standby distribution method.

Steps for configuring hot standby distribution

You can use hot standby distribution to have one preferred target server and one or more backup (hot standby) target servers. The preferred target server receives all new incoming connection requests and the hot standby servers act as backups in case the preferred target server becomes unavailable.

Before you begin

You need to determine which server should initially receive the workload (the preferred server), and which server or servers should be a backup server.

Procedure

Perform the following steps to configure hot standby distribution on the VIPADISTRIBUTE statement:

1. Configure HOTSTANDBY on the DISTMETHOD parameter.
2. Configure the AUTOSWITCHBACK parameter or the NOAUTOSWITCHBACK parameter.
   - Configure AUTOSWITCHBACK if you want the distributor to automatically switch distribution back to the preferred target when it is available. For example, if the preferred target becomes a standby target because its server is no longer ready, when the server is again in the listening state, the distributor automatically switches back to the preferred target as the active target. This is the default value.
   - Configure NOAUTOSWITCHBACK if you want the distributor to continue to use a backup target when the preferred target is available.

     Rule: If NOAUTOSWITCHBACK is configured, then the active server is initially determined by order of activation; the first ready listener becomes the active server regardless of the configured server type (PREFERRED or BACKUP), and remains the active server unless it becomes unavailable.
3. Configure the HEALTHSWITCH parameter or the NOHEALTHSWITCH parameter.
   - Configure HEALTHSWITCH if you want the distributor to automatically switch from the active target when it is not healthy. This is the default. A target is not healthy when there is a severe problem detected by one of the following health metrics:
     - Target server responsiveness (TSR) value is 0%
     - Rate of abnormal transaction completions is 1000
     - General health of the application is 0%

     For more information, see “Target server connection setup responsiveness monitoring” on page 468, and the WLM health metric information in “SERVERWLM - Distribution using WLM server-specific weights” on page 476.

     Rule: When a switch occurs because a target is no longer healthy, the target is not used as a backup target even if its health recovers, unless all available backup targets have previously had health problems. Health metric indicators typically recover when a server is no longer receiving any work, and although a server can appear healthy, the distributor cannot determine if the server is actually healthy. When the distributor detects that the server is transitioning back to the ready state, it again uses the target as a backup.

Tip: You can use the VARY TCPIP,,SYSPLEX,QUIESCE command and the VARY TCPIP,,SYSPLEX,RESUME command to manually bring a server back to the ready state. You can also use these commands before and after planned maintenance, or to temporarily divert new workload requests from a particular target. For more information, see "Manually quiescing DVIPA sysplex distributor server applications" on page 385.

- Configure NOHEALTHSWITCH if you want the distributor to ignore health metrics. The distributor switches from the active target only when it is not ready or when the distributor does not have an active route to the target.

4. Configure the PREFERRED option or the BACKUP option after each XCF address on the DESTIP parameter to designate the preferred server and the backup servers, and configure a rank value for each backup server.
   - Configure PREFERRED to designate the preferred server. This must be configured for one, and only one, XCF address.
   - Configure BACKUP to designate a backup (hot standby) server. This must be configured for at least one XCF address. Also configure a rank value on the BACKUP option for each backup server. The distributor switches to the highest ranked backup if the preferred server becomes unavailable.

Hot standby configuration example

The following example shows configuration for the hot standby distribution method:

```
VIPADISTRIBUTE DISTMETHOD HOTSTANDBY AUTOSWITCHBACK HEALTHSWITCH
   9.67.240.02 PORT 10000
   DESTIP
      203.3.10.16   PREFERRED
      203.3.10.17   BACKUP    50
      203.3.10.18   BACKUP   100
```

Because server 203.3.10.16 is the preferred server and AUTOSWITCHBACK is configured, that server becomes the active target when it enters the listening state. Servers 203.3.10.17 and 203.3.10.18 are the backup servers. The highest ranked backup server (203.3.10.18) becomes the active server if server 203.3.10.16 becomes unavailable.

Use the Netstat VIPADCFG/-F command to display the configuration:

```
Dynamic VIPA Information:
⋮
  VIPA Distribute:
    Dest:      9.67.240.02..10000
      DestXCF: 203.3.10.16
      DistMethod: HotStandby      SrvType: Preferred
      AutoSwitchBack: Yes         HealthSwitch: Yes
      SysPt:   No   TimAff: No    Flg:
      OptLoc:  No
    Dest:      9.67.240.02..10000
      DestXCF: 203.3.10.17
      DistMethod: HotStandby      SrvType: Backup  Rank: 050
      AutoSwitchBack: Yes         HealthSwitch: Yes
      SysPt:   No   TimAff: No    Flg:
      OptLoc:  No
    Dest:      9.67.240.02..10000
      DestXCF: 203.3.10.18
      DistMethod: HotStandby      SrvType: Backup  Rank: 100
      AutoSwitchBack: Yes         HealthSwitch: Yes
      SysPt:   No   TimAff: No    Flg:
      OptLoc:  No
```

Use the Netstat VDPT/-O command to display the status of the servers:

```
Dest:        201.2.10.15..5000
  DestXCF:   203.3.10.16
  TotalConn: 0000000000  Rdy: 001  WLM: 10 TSR: 100
  DistMethod: HotStandby           SrvType: Preferred
  Flg: Active
Dest:        201.2.10.15..5000
  DestXCF:   203.3.10.17
  TotalConn: 0000000000  Rdy: 001  WLM: 10 TSR: 100
```

```
  DistMethod: HotStandby           SrvType: Backup
  Flg: Backup
Dest:        201.2.10.15..5000
  DestXCF:   203.3.10.18
  TotalConn: 0000000000  Rdy: 001  WLM: 10 TSR: 100
  DistMethod: HotStandby           SrvType: Backup
  Flg: Backup
```

The Active and Backup flags show the current status of each server. For example, the preferred server (203.3.10.16) is displayed with the Active flag, indicating that it is the active server. Each server uses a default WLM weight of 10 for displays and health calculations.

If the active server becomes unavailable for some reason, the Netstat VDPT/-O command displays something similar to the following status:

```
Dest:        201.2.10.15..5000
  DestXCF:   203.3.10.16
  TotalConn: 0000000000  Rdy: 001  WLM: 00 TSR: 000
  DistMethod: HotStandby           SrvType: Preferred
  Flg: Backup
Dest:        201.2.10.15..5000
  DestXCF:   203.3.10.17
  TotalConn: 0000000000  Rdy: 001  WLM: 10 TSR: 100
  DistMethod: HotStandby           SrvType: Backup
  Flg: Backup
Dest:        201.2.10.15..5000
  DestXCF:   203.3.10.18
  TotalConn: 0000000000  Rdy: 001  WLM: 10 TSR: 100
  DistMethod: HotStandby           SrvType: Backup
  Flg: Active
```

The target server responsiveness (TSR) value for server 203.3.10.16 dropped to zero (TSR: 000), and the distributor switched to use the highest ranked backup, 203.3.10.18 (configured with a Backup rank of 100). The backup target becomes the active target (Flg: Active), and the preferred target becomes a backup target (Flg: Backup). However, because the switch from the preferred target occurred because the TSR value is 0, even when the TSR begins to recover, automatic switchback to this target does not occur. If server 203.3.10.16 is quiesced and resumed, the distributor can switch distribution back to this preferred server.

Timed affinity

The distribution method does not have any effect on incoming connection requests that have an active affinity established to a specific server instance (through the TIMEDAFFINITY parameter). When an affinity exists, it has priority over the distribution method setting.

SHAREPORT

Specifying the SHAREPORT parameter on the PORT statement in the TCP/IP profile enables a group of servers to listen on the same port, and thereby share the incoming workload. As new connections are received, the number of active connections is evenly balanced across the available servers using a weighted, round-robin distribution based on the Servers' accept Efficiency Fractions (SEFs). Specifying the SHAREPORTWLM parameter on the PORT statement enables connections to be distributed in a weighted, round-robin fashion based on the WLM server-specific recommendations modified by the SEFs. For more information on SEFs, see “Target server connection setup responsiveness monitoring” on page 468.

Use the same considerations concerning application type to determine the type of port sharing distribution to use. If the shared port is a sysplex distributed port and WLM server-specific weight is the distribution method that is being used by the distributor, code the SHAREPORTWLM paremeter on each target's PORT statement to take advantage of the WLM server-specific recommendations when connections are received at the target. Otherwise, use the SHAREPORT parameter on the PORT statement.

QDIO Accelerator

You can use QDIO Accelerator to provide accelerated forwarding at the DLC layer for some sysplex distributor packets. For more information, see "QDIO Accelerator" on page 101.

QDIO inbound workload queueing

You can use QDIO inbound workload queueing to improve throughput for inbound sysplex distributor packets. For more information, see "QDIO inbound workload queueing" on page 71.

Optimizing local connections

You can configure sysplex distributor to optimize connections when both connection endpoints potentially are on the same TCP/IP stack within the sysplex. This feature can be useful if your environment includes multi-tier server applications within a single sysplex. In this environment, if communication between the tiers is based on the TCP protocol, sysplex distributor can provide higher server availability. For example, without sysplex distributor or some other form of IP load balancing, when an application in tier 1 tries to connect to a tier 2 application on the local system, if the tier 2 server application is not available, these connections fail and the workload is disrupted. Sysplex distributor dynamically reroutes these connections to another tier 2 application that is running elsewhere in the sysplex, which provides a high availability solution. You can use the OPTLOCAL keyword to further optimize load balancing for this type of configuration; for more information about the OPTLOCAL keyword, see "Sysplex distribution optimizations for multi-tier z/OS workloads" on page 487.

Tip: In networks that have CISCO routers, it might be possible to remove the requirement that all inbound traffic needs to traverse the distributing stack.

Sysplex distributor also enhances the dynamic VIPA and automatic VIPA takeover functions. The enhancements allow a DVIPA to move nondisruptively to another stack. That is, in the past, a DVIPA was allowed to be active only on one single stack in the sysplex. This led to potential disruptions in service when connections existed on one stack, yet the intent was to move the DVIPA to another stack. With sysplex distributor, the movement of DVIPAs can now occur without disrupting existing connections on the original DVIPA owning stack.

See "Configuring distributed DVIPAs - sysplex distributor" on page 382 for more information.

Policy interactions

The Policy Agent interacts with the sysplex distributor to assist with workload balancing. There will be one Policy Agent running on an LPAR regardless of how many stacks are configured. First, the Policy Agent can be configured to collect network performance statistics for applications being distributed on target stacks. These network performance statistics are then used to modify the overall WLM weight assigned to a target server. In this way, processor performance, server performance, and application network performance are taken into account when distributing work. Second, policies established on the distributing stack can be configured to restrict the set of target stacks to be considered for any given inbound connection request. In this way, the total set of target stacks can be partitioned among different groups of users or applications requesting connections to distributed applications.

Previously, the QoS performance data was collected by the Policy Agent on the target for each DVIPA and port or application. After collecting the QoS information, the Policy Agent on the target stack pushed this information down to the stack sysplex function which then forwarded it to the stack sysplex function on the distributing stack. There are two significant additions to Policy Agent and sysplex interaction:

- The Policy Agent at each target will collect information with an additional level of granularity; the QoS performance data will be collected for each service level that a target's DVIPA port or application supports.
- The Policy Agent on the distributing stack drives the collection of this information by pulling it from the Policy Agents on the target stacks:
  - The Policy Agent on the distributor opens up one or two TCP connections to each of the Policy Agents on the target stacks. The distributor can be configured to distribute connections to IPv4 DVIPAs, IPv6

DVIPAs, or both. The Policy Agent opens an IPv4 connection to a target stack's IPv4 XCF address if it is configured to distribute to an IPv4 DVIPA on that target, and likewise opens an IPv6 connection to a target stack's IPv6 XCF address if it is configured to distribute to an IPv6 DVIPA on that target. As a result, if the routing stack is distributing connections to both IPv4 and IPv6 DVIPAs on a given target, then two connections to that target are opened.

For more information on how the sysplex distributor determines its targets, see "Configuring distributed DVIPAs - sysplex distributor" on page 382.

- The Policy Agent on the distributing stack will send across a list of QoS service level names to the Policy Agent on each target.
- The Policy Agent on each target will send back a QoS Policy Action weight fraction for each requested service level that each target DVIPA port/application supports. A specific Policy Action weight fraction will not be sent unless the distributing stack's Policy Agent requests it. Only weight fractions for IPv4 DVIPA ports and applications flow on the IPv4 Policy Agent connection, and similarly, only IPv6 weight fractions flow on the IPv6 Policy Agent connection.
- Upon receiving the QoS Policy Action weight fraction, the Policy Agent on the distributing stack will pass this information down to the sysplex distribution function on the stack. If two connections to a given target Policy Agent are active, weight fractions for IPv4 DVIPA ports or applications and IPv6 DVIPA ports or applications are passed to the sysplex distribution function separately. The stack sysplex distribution function uses this additional information when it is selecting targets for incoming connections. If it does not have a QoS Policy Action weight fraction, then it uses the existing weight fraction to make the load distribution decision instead.

Steps for enabling Policy Agent load distribution functions

The Policy Agent interacts with the sysplex distributor to assist with workload balancing.

Procedure

Perform the following steps to enable Policy Agent load distribution functions:

1. Define the PolicyPerfMonitorForSDR statement in the PAGENT configuration file to enable the policy performance monitor function.

   This function must be active on the target and distributing stacks.
2. z/OS Communications Server load distribution needs to be specifically enabled for each service level; a Policy Action with the same service level name needs to be defined on each of the appropriate target stacks and also on the distributing stack for these targets.

   Note that it is reasonable to have a subset of key service level names defined to the distributing stack. Traffic mapping to those service level names that are defined to the distributing stack will receive z/OS Communications Serverload distribution by service level. All other traffic will receive Communications Server V2R10 load distribution.
3. A backup distributing stack must have the same Policy Action configuration definitions as the active distributing stack for the corresponding DVIPA targets that it is backing up, if you want the Policy Action behavior to stay the same when the backup distributing stack takes ownership of the DVIPA. It also must have the Policy Agent performance monitor function active.
4. Common PAGENT port numbers will be used by the listener (pagentQosListener) and the collector (pagentQosCollector)

   They are part of the /etc/services install file. If PAGENT is running on an LPAR containing a target stack, it will open a listening connection using the pagentQosListener port number. PAGENT running on an LPAR containing a distributing stack will establish a TCP connection with each PAGENT listener using the pagentQosCollector as the source port and the pagentQosListener as the destination port. The listener will fail a connect request if the source/destination port does not match the defined collector/listener port. The /etc/services file on all LPARs in the sysplex must be updated to contain these port numbers.
5. Define these two port numbers as reserved ports for PAGENT using the PORT statement in the PROFILE.TCPIP data set.

6. Define the DYNAMICXCF parameter on the IPCONFIG or IPCONFIG6 statements in PROFILE.TCPIP. The PAGENT TCP connections use the XCF IP addresses.

Results

For more information on setting up polices, see “Sysplex distributor policy performance monitoring configuration” on page 834, “Sysplex distributor policy example” on page 840, or “Sysplex distributor routing policy example” on page 1395.

Sysplex distribution optimizations for multi-tier z/OS workloads

You can configure sysplex distributor to optimize connections when both connection endpoints potentially are on the same TCP/IP stack within the sysplex. This feature can be very useful in environments in which multi-tier server applications are located within the same z/OS system images in a single sysplex. In this environment, if communication between the tiers is based on the TCP protocol, you can use sysplex distributor to provide higher availability.

For example, without sysplex distributor or some other form of IP load balancing, if an application in tier 1 tries to connect to a tier 2 application on the local system, but the tier 2 server application is not available, then these connections fail and the workload is disrupted. With sysplex distributor, these connections are dynamically rerouted to another tier 2 application that is running elsewhere in the sysplex, which provides a high availability solution.

Examples of tier 1 server applications on z/OS that can be used with tier 2 server applications on the same z/OS logical partition (LPAR) are the IBM HTTP Server, the CICS Transaction Gateway, and the WebSphere Application Server. Examples of applications that can be used as tier 2 servers in the same z/OS system and sysplex as tier 1 servers are the CICS Transaction Server, IMS Connect, andDb2.

Sysplex distributor optimization with the OPTLOCAL keyword

Figure 56 on page 487 shows a sample configuration that uses the OPTLOCAL keyword.

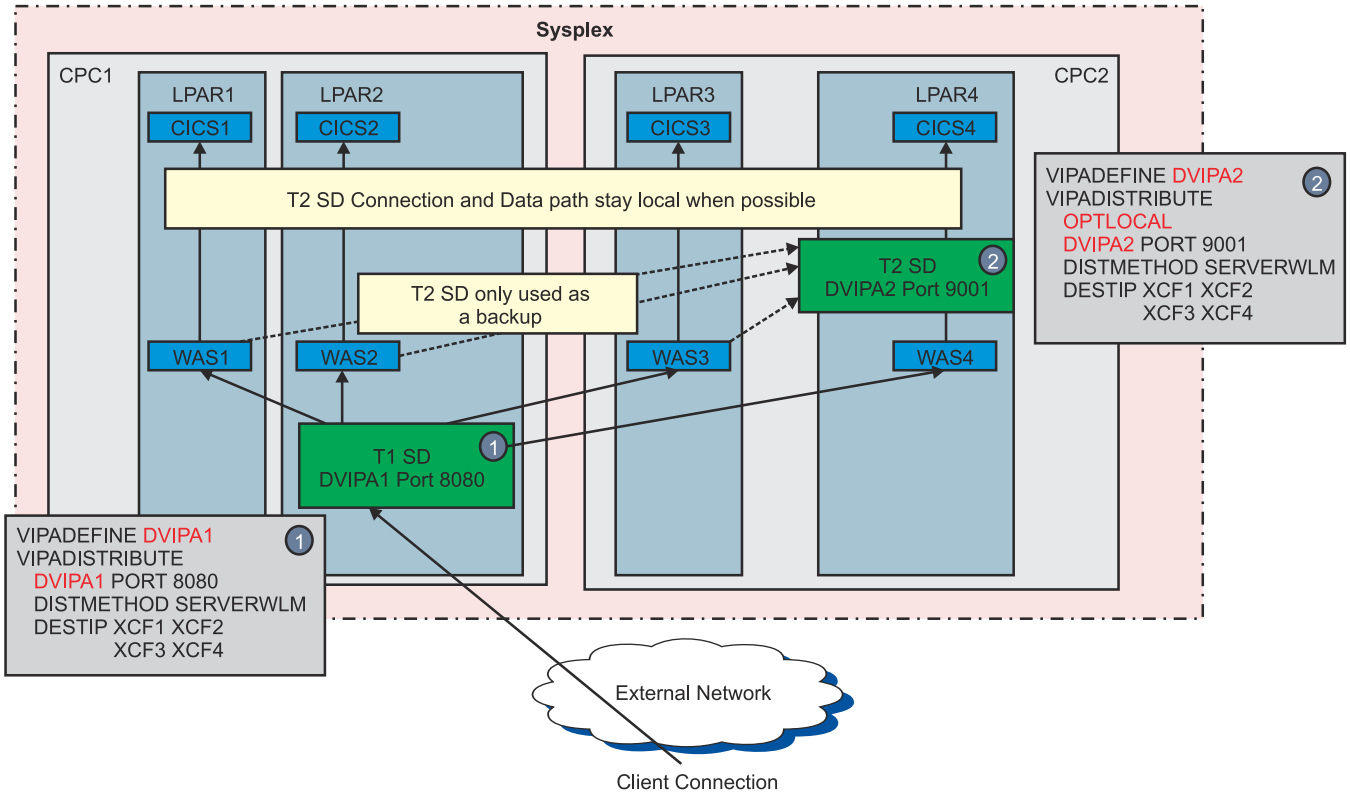


Figure 56. z/OS multi-tier application load balancing using sysplex distributor and the OPTLOCAL keyword

In Figure 56 on page 487, two tiers of servers are configured across four LPARs that are on two central processor complexes (CPCs) in a sysplex environment. The tier 1 servers consist of z/OS WebSphere Application Server (WAS) instances that are listening on port 8080 and that are represented by distributed DVIPA1. Sysplex distributor is configured on LPAR2 to receive and distribute any incoming connections for DVIPA1 to any of the active WAS instances in the four LPARs, based on server-specific WLM recommendations for each WAS instance (designated by the dynamic XCF address of each target LPAR). After a request is directed to a tier 1 WAS instance, the request is processed by that server. One or more secondary TCP connections can be established to a back-end tier 2 server, which is a cluster of CICS regions (one CICS region per LPAR).

You can use other sysplex distributor functions to distribute the tier 2 connections in this scenario, but that configuration does add some additional pathlength and processing, because all connection requests and traffic for those connections are forwarded through the sysplex distributor routing stack on LPAR4 so that they can get forwarded to the selected tier 2 server instance. This additional pathlength and processing occurs even in the case where the selected tier 2 server for a given connection is on the same LPAR as the tier 1 server that originated the connection.

To reduce excess pathlength but still retain high availability, you can specify the OPTLOCAL keyword for DVIPA2 to optimize processing in several ways:

- As long as local target resources are available and are not constrained, extra network flows and overhead through the distributor are avoided, because the local system makes the decision to route the connection to the local instance.
- When the decision is made to make the connection local, TCP/IP enables this connection for fast local sockets processing. This provides a more efficient path through the TCP/IP stack for these communications. All traffic for this connection remains on the local system and is not routed to the distributing stack.
- Optionally, multi-tier applications can also provide their own optimizations by using the Sysplex Sockets API (SO_CLUSTERCONNTYPE) or the trusted TCP connections API (SIOCGPARTNERINFO ioctl). These APIs enable the multi-tier applications to dynamically determine that they both are on the same system, and as a result, to optimize their processing based on their locality (for example, avoiding encryption of data, sharing memory, and so on). For more information about trusted TCP connections, see z/OS Communications Server: IP Programmer's Guide and Reference.
- If local target resources are unavailable or constrained, normal sysplex distributor processing is performed on the connections as a fallback, enabling the sysplex distributor to select another application that is available.

The OPTLOCAL feature provides for optimal performance in the most common scenario in which the local applications and systems are available and healthy, yet the feature also provides a high availability solution when the local applications or systems are not available or are overly constrained.

You can implement the OPTLOCAL feature using the OPTLOCAL keyword on the VIPADISTRIBUTE statement in the VIPADYNAMIC block, which enables a target stack to keep outbound connections for local resources on the local stack without having to send the connection request to the distributing stack. The level of preference that is shown to the local stack can be adjusted using the integer value specified on the OPTLOCAL keyword.

Value
Meaning

0
Specifying 0 means that the connections should remain local as long as the server is healthy. The relative capacity of the other systems is not considered in this case.

1
Specifying 1 is the same as specifying 0, except that if the WLM weight for the server on the local stack is 0, the connection request is forwarded to the sysplex distributor to find the best available server.

2 - 16
The values 2 through 16 are used as multipliers against the WLM weight for the server on the local stack, causing its weight to increase and therefore be favored over servers on other stacks. The larger the value specified, the more the local stack is favored.

If the OPTLOCAL keyword is specified with the value 0 or 1, the distribution method is not used for connections that originate from the specified target stack, if that target application is on the same stack and is able to handle its current workload. Regardless of the value that is specified on the OPTLOCAL keyword, connections are sent to the distributing stack if any of the following conditions are true:

- No local server is available.
- The server's accept efficiency fraction (SEF) has fallen below 75.
- The abnormal transaction completions value is greater than 250.
- The health indicator is less than 75.

For more information about the OPTLOCAL keyword on the VIPADISTRIBUTE statement in the VIPADYNAMIC block, see z/OS Communications Server: IP Configuration Reference.

Sysplex distributor enhanced workload distribution for z/OS multi-tier, OPTLOCAL configurations

In environments in which OPTLOCAL is used to optimize the load balancing between tier 1 and tier 2 server applications, more optimization is possible if sysplex distributor is also used as the load balancer for the tier 1 server applications. This optimization makes both tiers of z/OS server applications on a given system visible to sysplex distributor when making a load balancing decision on an incoming tier 1 connection request. When you are using WLM-based recommendations such as SERVERWLM, this optimization enables sysplex distributor to compute a composite WLM weight for each system, which includes the capacity, performance, and health characteristics of both the tier 1 server applications and the tier 2 server applications. Figure 57 on page 489 shows an example configuration that enables this enhancement.

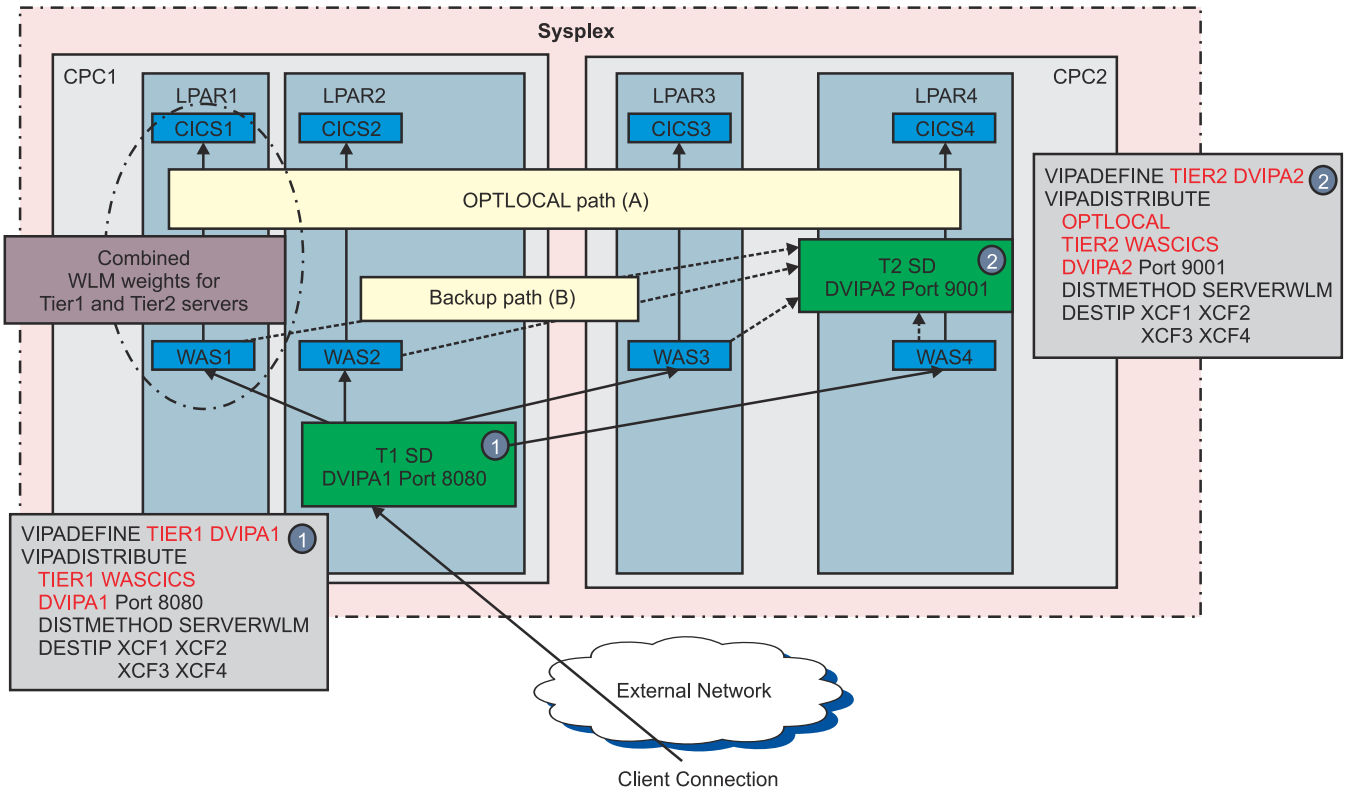


Figure 57. Enhanced z/OS multi-tier application load balancing using sysplex distributor

As shown in Figure 57 on page 489, two VIPADISTRIBUTE statements are required in this configuration. The first statement for DVIPA1 represents the distribution to the tier 1 server applications, a cluster of WebSphere Application Server instances running on each LPAR. The second statement for DVIPA2 represents the tier 2 server applications, a cluster of CICS regions on the same LPARs as the tier 1 servers.

The VIPADISTRIBUTE statement for each DVIPA includes a tier specification that indicates the role of each DVIPA (TIER1 or TIER2). The two DVIPAs are linked by a common group name specification, WASCICS. This association enables sysplex distributor to compute a single weight for each DVIPA1 target system that reflects the combination weight of the tier 1 and tier 2 server applications that are running on that system. Sysplex distributor performs its workload balancing of incoming connection requests to DVIPA1 based on these combined weights, enabling distribution of requests to the target z/OS systems to be based on the aggregate capacity of each system to perform the processing of both server application tiers.

In addition, because the VIPADISTRIBUTE statement for DVIPA2 also includes the OPTLOCAL specification, incoming tier 1 connection requests are typically directed to a system that also enables the subsequent tier 2 connection requests to stay local. This provides an improvement over the base OPTLOCAL configuration, because new workload is directed toward systems that have more capacity to process the work while retaining the local optimizations provided by OPTLOCAL.

Rule: To use this feature, the distribution method for the tier 1 and tier 2 servers must be BASEWLM or SERVERWLM.

Guideline: To take full advantage of this feature, set the distribution method for both tier 1 and tier 2 servers to SERVERWLM. This enables sysplex distributor to obtain server-specific WLM recommendations for each application server tier; the combined weight reflects the composite performance of the servers instead of the capacity of the system.

For more information about the TIER1 and TIER2 keywords on the VIPADISTRIBUTE statement in the VIPADYNAMIC block, see z/OS Communications Server: IP Configuration Reference.

Sysplex distributor enhanced workload distribution for z/OS multi-tier, OPTLOCAL configurations with CPC affinity

In addition to multi-tier application awareness and optimizing distribution when these applications are local, you can use the CPCSCOPE keyword to further optimize sysplex distributor load balancing decisions. This configuration option helps to avoid workload distribution to tier 1 servers that cannot perform the tier 2 processing on the same CPC, which requires distribution of those tier 2 connections to systems that are on a different CPC. Figure 58 on page 491 shows an example configuration in which this option might be useful.

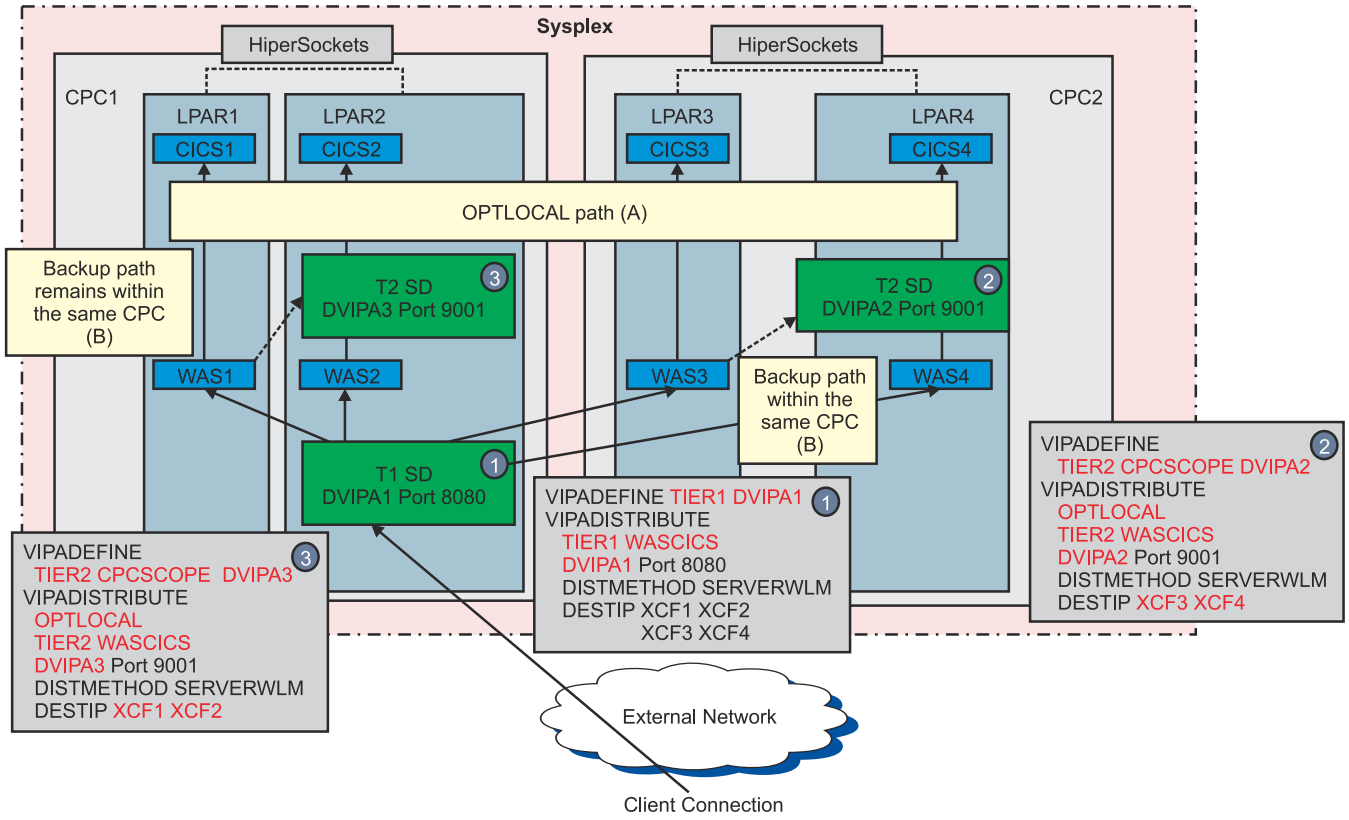


Figure 58. z/OS multi-tier application configuration using CPCSCOPE DVIPAs

In Figure 58 on page 491, connection requests for tier 1 server applications are distributed to systems based on the aggregate WLM recommendations of the tier 1 and tier 2 server applications running on each system. When connections are directed to a tier 1 server instance, any subsequent tier 2 connections that are originated by that server instance are likely to be directed to the local tier 2 server as a result of the OPTLOCAL keyword, assuming that the tier 2 server application is active on the local system.

If the tier 1 server cannot access a tier 2 server on the same system, the connection requests and their associated data flows are sent to the sysplex distributor for the tier 2 DVIPA; this distributor might be on another system in the sysplex. That system might be on the same CPC as the tier 1 server, or it might be on another CPC at the same site or potentially at a remote site. The sysplex distributor then selects a target tier 2 server instance that is on the same CPC or in another CPC. In either case, inbound traffic from the tier 1 server to the tier 2 server might need to traverse one or more CPCs or external network links before it reaches the tier 2 server.

You can optimize these flows by using the CPCSCOPE keyword and confining the activation and movement of a DVIPA to a specific CPC, which in turn enables any connection using the backup path to a tier 2 server to also remain within an LPAR on the same CPC. This optimization enables these backup data flows to leverage secure, high speed, virtual network links inside the CPC, such as HiperSockets. This configuration also helps to ensure that these data flows are optimized by avoiding the need to traverse external links, which might result in additional network latency and encryption requirements.

Three DVIPAs are defined in Figure 58 on page 491:

- DVIPA1 represents the tier 1 servers that are on LPAR1 - LPAR4. DVIPA1 is defined as a tier 1 DVIPA, and is associated with the group name WASCICS.
- DVIPA2 represents the tier 2 servers that are on LPAR3 and LPAR4 on CPC2. DVIPA2 is defined as a tier 2 DVIPA, and is associated with the group name WASCICS. DVIPA2 also specifies the CPCSCOPE keyword, which ensures that DVIPA2 remains within LPARs on CPC2, even in recovery scenarios where DVIPA takeover is required.
- DVIPA3 represents the tier 2 servers that are on LPAR1 and LPAR2 on CPC1. DVIPA3 is similar to DVIPA2; it is also a tier 2 DVIPA, is associated with the same group name (WASCICS), and also specifies

the CPCSCOPE keyword. If a takeover is required for this DVIPA, it remains within any backup LPARs on CPC1.

Requirement: Configure the tier 1 application servers on each CPC to use a unique, CPC-specific DVIPA for their tier 2 connections. The tier 1 application server configurations need to be unique based on the CPC where they are.

Guideline: Before activating this feature, consider the benefits that the feature provides versus the additional configuration requirements that this type of configuration requires. You must carefully define the tier 2 DVIPAs to ensure that the correct VIPADEFINE and VIPBACKUP definitions are maintained on each of the LPARs based on the CPC on which the LPARs are defined. These definitions can have implications on your recovery plans and procedures that might allow z/OS systems or tier 1 server applications to be restarted on LPARs in different CPCs, rather than in the CPC in which they were originally activated.

For more information about the CPCSCOPE keyword on the VIPADISTRIBUTE statement in the VIPADYNAMIC block, see z/OS Communications Server: IP Configuration Reference.

Chapter 9. TCP/IP in an ensemble

OSA-Express Ethernet features in QDIO mode can access the external data network when configured with the OSD channel path ID (CHPID) type. The OSD CHPID type is the default CHPID type on the IPAQENET and IPAQENET6 INTERFACE statements in the TCP/IP profile.

IBM zEnterprise System (zEnterprise) provides communications access to two internal networks through OSA-Express3 or later adapters that are configured with the OSA-Express for Unified Resource Manager (OSM) or OSA-Express for zBX (OSX) CHPID types. Communications Server supports OSA-Express3 or later adapters that are configured with these CHPID types, thus allowing TCP/IP connectivity to the following internal networks.

- Intraensemble data network (CHPID type OSX)

  The intraensemble data network (IEDN) provides access to other images connected to the IEDN, and to applications and appliances that are running in an IBM zEnterprise BladeCenter Extension (zBX). The IEDN can be accessed through 10 gigabit OSA-Express3 or later adapters that are configured with CHPID type OSX. OSX interfaces can be IPv4 or IPv6, and must be on a VLAN. To configure OSX interfaces, see "Steps for configuring an interface for the intraensemble data network (CHPID type OSX)" on page 494.

  OSX connectivity also enables HiperSockets connectivity to the IEDN; you configure an Internal Queued Direct I/O (IQD) CHPID for HiperSockets to enable the Internal Queued Direct I/O extensions (IQDX) function, referred to as the z/OS Communications Server IEDN-enabled HiperSockets function. For more information about the z/OS Communications Server IEDN-enabled HiperSockets function and its benefits, see "HiperSockets connectivity to the intraensemble data network" on page 494.
- Intranode management network (CHPID type OSM)

  The intranode management network (INMN) is an IPv6 network that provides connectivity between network management applications within a zEnterprise node, and can be accessed through 1000BASE-T Ethernet OSA-Express3 or later adapters that are configured with CHPID type OSM. OSM interfaces are not configured in the TCP/IP profile, but are generated by the stack and have only link-local IP addresses. The stack does not report OSM interfaces to OMPROUTE, and OMPROUTE is unaware of these interfaces; there are no OMPROUTE configuration considerations for OSM interfaces. To use the INMN, see "Steps for using the intranode management network (CHPID type OSM)" on page 499.

Requirement: To access the IEDN or the INMN, the IBM zEnterprise 196 (z196) central processor complex (CPC) and the logical partition (LPAR) must be configured as members of an ensemble. For information about using the ENSEMBLE start option to specify that an LPAR is a member of an ensemble, see z/OS Communications Server: SNA Resource Definition Reference.

For more information about zEnterprise and ensembles, see the following information in IBM Documentation at https://www.ibm.com/docs/en:

- *zEnterprise BladeCenter Extension Installation Manual*
- *zEnterprise System Introduction to Ensembles*
- *IBM z Systems Ensemble Workload Resource Group Management Guide*
- *IBM z Systems Ensemble Planning Guide*

© Copyright IBM Corp. 2000, 2022

Steps for configuring an interface for the intraensemble data network (CHPID type OSX)

The intraensemble data network (IEDN) can be accessed through OSA-Express for zBX (OSX) interfaces. The IEDN provides access to other images connected to the IEDN, and to applications and appliances running in an IBM zEnterprise BladeCenter Extension (zBX).

Before you begin

- To access the IEDN, the IBM zEnterprise 196 (z196) central processor complex (CPC) and the logical partition (LPAR) must be configured as members of an ensemble.

  For information about using the ENSEMBLE start option to specify that an LPAR is a member of an ensemble (ENSEMBLE=YES), see z/OS Communications Server: SNA Resource Definition Reference.
- Before the OSX device can become active, the Hardware Management Console (HMC) must be configured in the following way:
  - The VLAN ID specified at the HMC must match the ID specified on the VLANID parameter of the IPAQENET or IPAQENET6 statement.
  - The HMC must have added the LPAR and CHPID combination to the definition for that VLAN.

  See the information about creating and managing virtual server networks in *IBM z Systems Ensemble Planning Guide*.
- To use a zBX with OSX, you must configure the zBX from the HMC to use addresses on the same network specified on the IPAQENET or IPAQENET6 statements.

  See the information about obtaining the pairing code for accelerator authentication in the *IBM Smart Analytics Optimizer for Db2 for z/OS: Installation Guide*.

Procedure

Perform the following steps to configure an interface for the IEDN:

1. Specify CHPIDTYPE OSX on the INTERFACE statement for the IPAQENET or IPAQENET6 interface.
2. Include the CHPID parameter to have VTAM dynamically create the associated transport resource list element (TRLE) definition, and include the VLANID parameter.

   Unlike OSD interfaces, OSX interfaces must be associated with a specific VLAN, and can communicate only with other applications and images that have access to the same VLAN on the IEDN.

   The following example is a sample INTERFACE definition; adjust values as needed:

   ```
   INTERFACE QDIOOSX1 DEFINE IPAQENET
   CHPIDTYPE OSX    CHPID F1    VLANID 10
     MTU 8992     IPADDR 172.16.1.1/24
   ```
3. For IPv6, if you connect an external router on this network, then you can use stateless address autoconfiguration. Otherwise, you must configure IPv6 addresses and prefixes manually.

Results

For more information about configuring CHPID type OSX for IPAQENET and IPAQENET6 interfaces, see z/OS Communications Server: IP Configuration Reference. Some INTERFACE statement defaults are different for OSX interfaces than for OSD interfaces, and some INTERFACE statement parameters that apply to OSD interfaces do not apply to OSX interfaces.

HiperSockets connectivity to the intraensemble data network

IBM zEnterprise System (zEnterprise) provides the capability to integrate HiperSockets connectivity to the intraensemble data network (IEDN); HiperSockets connectivity to the IEDN is referred to as the z/OS Communications Server IEDN-enabled HiperSockets function. Within each central processor complex

(CPC) that is a member of an ensemble, you can define a single Internal Queued Direct I/O (IQD) channel path ID (CHPID) for HiperSockets to provide connectivity to the IEDN; you configure the designated IQD CHPID using a channel parameter in the hardware configuration definition (HCD), which enables the Internal Queued Direct I/O extensions (IQDX) function of HiperSockets. The IQDX function is a channel function and is not a new CHPID type. When the IQDX function is configured, the single IQD CHPID is integrated with the IEDN; for communications within the CPC, the IQDX function enables secure network access using the designated IQD CHPID.

The z/OS Communications Server IEDN-enabled HiperSockets function provides the following benefits:

- Combines the existing high-performance attributes of HiperSockets for intra-CPC communications with the secure access control, virtualization, and management functions of the IEDN that are provided by zEnterprise Unified Resource Manager (zManager)
- Converges OSA connectivity and HiperSockets connectivity into a single logical network interface

  The single interface simplifies network management by reducing the number of resources that are required to represent individual unique networks, including the following resources:

  – IQD CHPIDs
  – IP interfaces
  – IP subnetworks (VLANs)
  – IP addresses
  – IP routes
- Eliminates the HiperSockets configuration tasks within z/OS Communications Server and zManager

  The HiperSockets network configuration is inherited from the OSA-Express for zBX (OSX) configuration.
- Simplifies z/OS movement by eliminating or minimizing reconfiguration tasks

  The z/OS network configuration that is related to the IP topology (IP address, IP subnet, and VLAN) for the IEDN, which is configured with the OSX interface, does not change when a z/OS image is moved to a different CPC. Given that IQDX configuration attributes are inherited from the OSX configuration, no reconfiguration tasks are required for IQDX interfaces.
- Enables sysplex network traffic to transparently use HiperSockets connectivity for VIPAROUTE processing over OSX interfaces

Figure 59 on page 496 shows the key concepts of the z/OS Communications Server IEDN-enabled HiperSockets function; HiperSockets connectivity is transparently converged with OSA-Express connectivity.

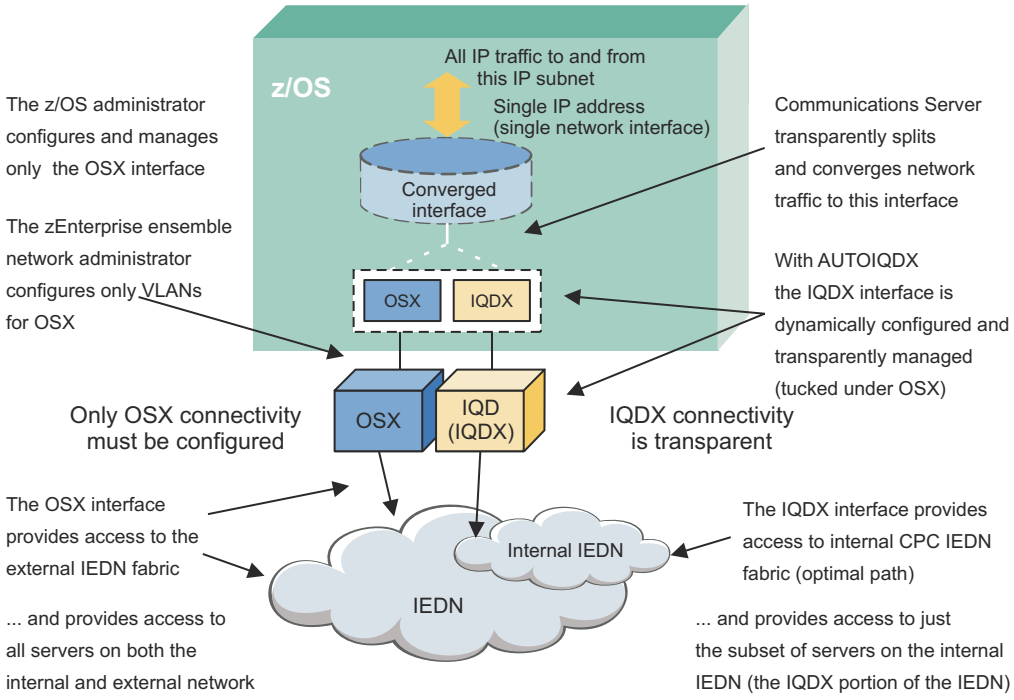


Figure 59. z/OS Communications Server IEDN-enabled HiperSockets overview

Communications Server provides transparent access to the IQD CHPID that is configured for the IQDX function. For each configured OSX CHPID and for each IP version, Communications Server dynamically creates and associates an IQDX interface and transport resource list element (TRLE) to the IQD CHPID. The dynamically created IQDX interfaces inherit the VLAN IDs and IP addresses of their associated OSX interfaces, thus eliminating the need for IP topology, routing, or security changes in the network. Communications Server dynamically determines whether traffic routed to the IEDN over a given OSX CHPID can also use the IQDX interface; if the traffic can use the IQDX interface, Communications Server routes the traffic over the high-performance HiperSockets IQD CHPID.

IQDX interfaces and TRLEs are created when the AUTOIQDX parameter is specified (or is specified by default) on the GLOBALCONFIG statement in the TCP/IP profile and an IQD CHPID is configured for the IQDX function.

- The naming convention for the dynamically created IQDX interfaces is EZAIQX*xx* (IPv4) and EZ6IQX*xx* (IPv6), where *xx* is the OSX CHPID number to which a particular IQDX interface is associated.

  The IQDX interfaces are identified by interface type IPAQIQDX for IPv4 and interface type IPAQIQDX6 for IPv6. Traffic for unique VLANs over the same OSX CHPID (multiple VLANs) share the same IQDX interface.
- The naming convention for the dynamically created IQDX TRLEs is IUTIQX*xx* (IPv4) and IUTIQ6*xx* (IPv6), where *xx* is the OSX CHPID number to which a particular IQDX interface is associated.

To enable Communications Server to access the IEDN using HiperSockets interfaces, see "Steps for enabling HiperSockets access to the intraensemble data network" on page 498.

Operating and managing IEDN-enabled HiperSockets interfaces

Internal Queued Direct I/O extensions (IQDX) interfaces are dynamically operated. They are started and stopped when the associated OSA-Express for zBX (OSX) interfaces are started and stopped. Optionally, IQDX interfaces can be separately operated if at least one of the associated OSX interfaces is active. If an IQDX interface fails, the TCP/IP stack dynamically attempts recovery, as it does with other configured HiperSockets devices and interfaces.

z/OS Communications Server dynamically allocates two control channels and up to eight data devices for each dynamically created IQDX transport resource list element (TRLE). TCP/IP stacks that share OSX CHPIDs also share dynamically created IQDX TRLEs; each stack uses one of the data devices for the IQDX

TRLE. You should configure sufficient subchannels for the Internal Queued Direct I/O (IQD) CHPID that is configured for the IQDX function. For more information about defining IQD CHPIDs and device types, see z/OS HCD Planning.

Routing, IP security, and packet tracing of the dynamic IQDX interfaces is controlled through their associated OSX interfaces; the IQDX interfaces are generally not visible to these functions.

- Routing daemons such as OMPROUTE are aware of only the associated OSX interfaces; the stack does not report IQDX interfaces to OMPROUTE, and static routes cannot reference the dynamic IQDX interfaces. z/OS Communications Server makes the decision to direct traffic over an IQDX interface after the associated OSX interface is chosen as the route for the traffic.
- IQDX interfaces are not directly visible to IP security. If IP security is enabled, then traffic over an IQDX interface inherits the security class (SECCLASS) from the associated OSX interface. For packets that are sent and received over the IQDX interfaces, IPSec filtering applies to the security class of the associated OSX interface.
- Packet tracing of an OSX interface includes traffic that is redirected or received over its associated IQDX interface. If you perform a packet trace for data over an IQDX interface, the trace includes only inbound packets that are discarded before the stack can determine the associated OSX interface.

You can use Netstat reports to display information about dynamically created IQDX interfaces.

- The Netstat DEvlinks/-d report displays information for each IQDX interface, as it does for any defined interface. The Netstat DEvlinks/-d report for a given OSX interface displays how much traffic was rerouted to the associated IQDX interface. For more information about the Netstat DEvlinks/-d report, see z/OS Communications Server: IP System Administrator's Commands.
- You can monitor which IP addresses are reachable over the OSX interfaces and over the associated IQDX interfaces using the Netstat ARp/-R and Netstat ND/-n reports. For more information about the Netstat ARp/-R report and the Netstat ND/-n report, see z/OS Communications Server: IP System Administrator's Commands.
- You can display information about the dynamic IQDX TRLEs and their associated datapath devices by issuing the D NET,ID=*trle* command or the D NET,TRL,TRLE=*trle* command. For more information about the DISPLAY ID command and the DISPLAY TRL command, see z/OS Communications Server: SNA Operation.

Performance considerations for the IEDN-enabled HiperSockets function

For network traffic within a central processor complex (CPC), HiperSockets processing can provide a fast response time (low latency). This fast response time is most likely for interactive (request and response) workloads where low latency is critical. However, when large amounts of data are transmitted (typically, streaming or bulk data workloads) as a significant portion of the intra-CPC workload, you might prefer OSA-Express processing due to differences in the OSA-Express processing for large data transmissions. The savings offered by OSA-Express processing for large data transfers vary; HiperSockets transmissions generally use more processor cycles than OSA transmissions, and this difference in processing cycles is magnified as transmission size grows.

If you prefer to use OSA-Express processing for large transmissions, the Communications Server IEDN-enabled HiperSockets function uses the strengths of both the HiperSockets and OSA-Express technologies, providing a configuration option that transparently controls traffic selection. Specifying NOLARGEDATA on the AUTOIQDX parameter of the GLOBALCONFIG statement transparently directs large data transmissions over OSA-Express for zBX (OSX) interfaces while directing small data transmissions over Internal Queued Direct I/O extensions (IQDX) interfaces; you can enable this setting to get the fast HiperSockets response time for small data transmissions, while avoiding the processor cost penalty sometimes incurred with HiperSockets for large data transmissions. If you do not anticipate a significant amount of large data transmissions or you always need high throughput levels (regardless of processor cycle consumption), you should not specify NOLARGEDATA on the AUTOIQDX parameter; if you do not specify NOLARGEDATA on the AUTOIQDX parameter, IQDX interfaces are favored for all transmission sizes.

Guidelines:

- When NOLARGEDATA is specified for the AUTOIQDX parameter, Communications Server forces large TCP transmissions to the OSX path. Communications Server defines a large transmission as a TCP socket send of 32 K or larger. All other traffic is transmitted over IQDX interfaces.
- When an OSA adapter is shared among two or more logical partitions (LPARs), the processor savings offered by the OSA adapter is less than the savings offered by configurations that use multiple OSA adapters that are not shared. Some OSA assist functions, such as segmentation offload, are not available for the shared OSA configuration.
- Regardless of whether NOLARGEDATA is specified for the AUTOIQDX parameter, if large data workloads are part of the overall IEDN workload, then you should use the HiperSockets multiple write assist with IBM Z® Integrated Information Processor (zIIP) function, which minimizes the increase in general-purpose processor usage that is incurred with HiperSockets. You should also use the 64 K Internal Queued Direct I/O (IQD) frame size.
- To be sure to take advantage of IQDX for streaming workloads, use an IQDX frame size larger than 16K if you are using an MTU size larger than 8192 bytes for OSX. Otherwise, certain streaming workloads are not able to use IQDX.

Steps for enabling HiperSockets access to the intraensemble data network

OSA-Express for zBX (OSX) connectivity enables HiperSockets connectivity to the intraensemble data network (IEDN). You configure an Internal Queued Direct I/O (IQD) channel path ID (CHPID) for HiperSockets with a channel parameter to enable the Internal Queued Direct I/O extensions (IQDX) function, referred to as the z/OS Communications Server IEDN-enabled HiperSockets function.

Before you begin

- To access the IEDN, the IBM zEnterprise 196 (z196) central processor complex (CPC) and the logical partition (LPAR) must be configured as members of an ensemble.

  For information about using the ENSEMBLE start option to specify that an LPAR is a member of an ensemble (ENSEMBLE=YES), see z/OS Communications Server: SNA Resource Definition Reference.
- Configure all OSX interfaces to the IEDN.

  OSX connectivity to the IEDN is required to enable HiperSockets connectivity to the IEDN. For OSX configuration information, see “Steps for configuring an interface for the intraensemble data network (CHPID type OSX)” on page 494.
- Define a HiperSockets CHPID for the IQDX function using a channel parameter in the hardware configuration definition (HCD), and sufficient subchannel addresses to enable creation of one dynamic IQDX transport resource list element (TRLE) per OSX CHPID per IP version.

  You must define 10 IQDX subchannel addresses for each OSX CHPID that is in use for IPv4, and 10 subchannel addresses for each OSX CHPID that is in use for IPv6. Multiple VLAN does not affect the required number of subchannel addresses. For more information, see the information about defining IQD CHPIDs and device types in z/OS HCD Planning.

Procedure

Perform the following steps to enable HiperSockets access to the IEDN.

1. Determine whether you want large outbound TCP protocol data to be transported to the IEDN over OSX interfaces or IQDX HiperSockets interfaces.

   For more information, see “Performance considerations for the IEDN-enabled HiperSockets function” on page 497.
2. If necessary, configure the appropriate value for the AUTOIQDX parameter on the GLOBALCONFIG statement in the TCP/IP profile.

- If you want all eligible traffic to be transported over IQDX HiperSockets interfaces, no other action is necessary. The AUTOIQDX parameter with the value ALLTRAFFIC is the default setting on the GLOBALCONFIG statement in the TCP/IP profile.
- If you want large outbound TCP protocol data transmissions to be transported to the IEDN over OSX interfaces, specify the value NOLARGEDATA on the AUTOIQDX parameter of the GLOBALCONFIG statement in the TCP/IP profile.

Steps for enabling IPv6 on a stack for access to the intranode management network

The intranode management network (INMN) is an IPv6 network that provides connectivity between network management applications within a zEnterprise node

Procedure

Perform the following steps to enable IPv6 on a stack so that the stack can access the INMN:

1. Use the NETSTAT CONFIG command to view the current TCP/IP configuration.

   If an `IPv6 Configuration Table` section is reported, no further actions are required.
2. Test and migrate any scripts that you use to format Netstat output in long format.

   After you enable a z/OS TCP/IP stack for IPv6, Netstat output for the stack is available in the long format only. While you are still using IPv4 for your data networks, you can test the Netstat reports in the long format using one of the following methods:

   - Use the FORMAT/-M LONG option on the Netstat command.
   - Specify the FORMAT LONG parameter on the IPCONFIG statement in your TCP/IP profile.

   For more information about the Netstat command, see z/OS Communications Server: IP System Administrator's Commands. For more information about the IPCONFIG statement, see z/OS Communications Server: IP Configuration Reference.
3. Update the BPXPRM*xx* parmlib member to include an additional NETWORK statement for IPv6.

   Find the existing NETWORK DOMAINNAME(AF_INET) statement, note the MAXSOCKETS and TYPE values that are specified, and add the following statement after the existing statement using those same values:

   ```
   NETWORK DOMAINNAME(AF_INET6)
   DOMAINNUMBER(19)
   MAXSOCKETS(number)
   TYPE(file_system_type)
   ```

   For more information about updating BPXPRM*xx*, see “Defining TCP/IP as a UNIX System Services physical file system” on page 39.

Steps for using the intranode management network (CHPID type OSM)

The intranode management network (INMN) is an IPv6 network that provides connectivity between network management applications within a zEnterprise node. The INMN can be accessed through 1000BASE-T Ethernet OSA-Express3 or later adapters that are configured with channel path ID (CHPID) type OSM.

Before you begin

The stack must be enabled for IPv6. For information about enabling a stack for IPv6, see “Steps for enabling IPv6 on a stack for access to the intranode management network” on page 499.

To access the INMN, the IBM zEnterprise 196 (z196) central processor complex (CPC) and the logical partition (LPAR) must be configured as members of an ensemble. For information about using the

ENSEMBLE start option to specify that an LPAR is a member of an ensemble, see z/OS Communications Server: SNA Resource Definition Reference.

If the stack is enabled for IPv6 and the LPAR is configured as a member of an ensemble, Communications Server automatically configures, locates, and activates up to two interfaces (named EZ6OSM01 and EZ6OSM02) onto the INMN. Each of these interfaces is an IPv6 interface that has only a link-local IP address. You cannot configure static or dynamic routes over OSA-Express for Unified Resource Manager (OSM) interfaces.

The INMN is intended for only authorized applications, such as those performing platform Performance Management functions. For more information about these applications, see *IBM z Systems Ensemble Planning Guide*.

Procedure

Perform the following steps to use the INMN:

1. Authorize the application to the EZB.OSM.*sysname*.*tcpname* resource.

   To use IOCTL calls to retrieve information about an OSM interface or to send or receive data over an OSM interface, an application must have READ authorization to the EZB.OSM.*sysname*.*tcpname* resource. If used on this image, authorize the application to this resource.
2. Reserve the UDP port that the platform management application is to use to listen for multicast traffic over the INMN.
3. Authorize any user IDs to this resource that might issue diagnostic commands, such as Ping and Traceroute, over OSM interfaces to verify connectivity.
4. If you enable IP security for IPv6, you can configure a security class for IP filtering that applies to all OSM interfaces.

   Use the OSMSECCLASS parameter on the IPCONFIG6 statement. This enables you to configure filter rules for traffic over the EZ6OSM01 and EZ6OSM02 interfaces.
5. If the multicast address that is used by the platform management application is configured into a network access zone, then give the user ID for this application read permission to the resource profile for that zone.

Routing considerations for the intraensemble data network

The TCP/IP stack does not allow IP forwarding within the intraensemble data network (IEDN) between two OSA-Express for zBX (OSX) interfaces. However, the stack does allow IP forwarding within the IEDN by the sysplex distributor when VIPAROUTE is used. The stack also allows IP forwarding between an OSX interface and a non-OSX interface.

If an external router is attached to the IEDN, dynamic and static routing considerations for OSX interfaces are the same as those considerations for OSD interfaces connected to an external router. However, if no external router is present on that network, see the following guidelines.

Guidelines:

- Implement the IEDN as a flat network, meaning that all addresses that are reachable through the IEDN are in the same subnet (or IPv6 prefix) as the IEDN. In this configuration, a single route (static or dynamic) to this subnet enables all addresses on the IEDN to be reached.
- Static routing might be appropriate in the following cases:
  - The IEDN is implemented as a flat network, in which case a single static route over the OSX interface to the subnet of the IEDN can reach all destinations in the IEDN.
  - A z/OS host will not be routing traffic between external networks and the IEDN, so you do not want to advertise the intraensemble subnet addresses to the external network.

    Tip: If you are using static routing over the IEDN and there are destination IP addresses on the IBM zEnterprise BladeCenter Extension (zBX) that are not part of the intraensemble subnet, you must

define static routes to those destination IP addresses on the intraensemble network by using OSX interfaces.

- Dynamic routing might be appropriate in the following cases:
  - You want destinations on the intraensemble subnet to be reachable from outside the ensemble, and therefore want to advertise the IEDN addresses to the external network.
  - There are destination IP addresses on the zBX that are not part of the intraensemble subnet, and you do not want to code individual static routes to reach them (for example, if other hosts on the IEDN have VIPA addresses that are not in the intraensemble subnet).

OMPROUTE considerations for the intraensemble data network

In an ensemble environment with dynamic routing where a z/OS stack is the router for the ensemble, you can control whether the subnet of the intraensemble data network (IEDN) is advertised.

- If you want to prevent external traffic from being routed to this VLAN, then do one of the following items so that OMPROUTE does not advertise the intraensemble subnet:
  - Define the OSA-Express for zBX (OSX) interface to OMPROUTE using an INTERFACE statement or IPV6_INTERFACE statement, and do not enable the IMPORT_DIRECT_ROUTES function of AS boundary routing. This function is disabled by default, so ensure that the IMPORT_DIRECT_ROUTES parameter of the AS_BOUNDARY_ROUTING statement or the IPV6_AS_BOUNDARY_ROUTING statement is not set to YES.
  - Do not define the OSX interface to OMPROUTE, and ensure that GLOBAL_OPTIONS IGNORE_UNDEFINED_INTERFACES is configured to OMPROUTE.

  You might also want to install IPSec filter rules to restrict which traffic can flow on this network.
- If you want to allow external traffic to be routed to this VLAN, then define the OSX interface to OMPROUTE as an OSPF_INTERFACE or IPV6_OSPF_INTERFACE, and code a nonzero value for the ROUTER_PRIORITY parameter on the interface. As long as no other hosts on that OSX VLAN have coded their interfaces as OSPF interfaces, then OMPROUTE advertises the subnet (or IPv6 prefixes) of the IEDN into the OSPF network. This advertisement makes all addresses that fall into the intraensemble subnet (or IPv6 prefixes) reachable using OSPF. In addition, if you have a subset of hosts on the IEDN with VIPAs that are not in the intraensemble subnet, which you want to be able to reach over the IEDN, you can define the OSX interfaces on those hosts as OSPF interfaces; OMPROUTE communicates with them (and only them) and can route to their VIPAs, while still preserving the ability to route to the other hosts by subnet only.

Tip: These definitions apply per interface, so you could implement advertising on one VLAN while not advertising on a different VLAN attached to the same z/OS router.

Sysplex distributor considerations for the intraensemble data network

In a sysplex distributor configuration, if the client, sysplex distributor, and at least one sysplex distributor target are all connected over the intraensemble data network (IEDN), then the client, sysplex distributor, and all potential sysplex distributor targets must be connected to the same VLAN on the IEDN.

Multilevel security and network access control considerations

In a multilevel secure environment, you should treat the intraensemble data network (IEDN) and the OSA-Express for zBX (OSX) interfaces as any other data network with OSA-Express access. Because access to the IEDN through dynamic Internal Queued Direct I/O extensions (IQDX) interfaces is dependent on OSX access, traffic that is permitted or restricted over OSX interfaces is similarly permitted or restricted over IQDX interfaces. However, the intranode management network (INMN) and the OSA-Express for Unified Resource Manager (OSM) interfaces require special considerations. For information about these additional considerations, see Chapter 4, “Preparing for IP networking in a multilevel secure environment,” on page 191.

If you are using network access control, the IEDN and the OSX interfaces are subject to the same network access control as any other data network with OSA-Express access. Because access to the IEDN through dynamic IQDX interfaces is dependent on OSX access, traffic that is permitted or restricted over OSX interfaces is similarly permitted or restricted over IQDX interfaces. However, all traffic to and from the INMN over OSM interfaces is exempt from network access control, and is instead subject to OSM access control. Only multicast addresses to which a platform management application binds are subject to network access control.

Chapter 10. Shared Memory Communications

Shared Memory Communications (SMC) enables two SMC capable peers to communicate by using memory buffers that each peer allocates for the partner's use. There are two types of Shared Memory Communications:

- Shared Memory Communications over Remote Direct Memory Access (SMC-R)
- Shared Memory Communications - Direct Memory Access (SMC-D)

SMC improves throughput, lowers latency and cost, and maintains existing functions. You do not need to change resources, such as host names and IP addresses, because you can use existing IP topology and addressing to identify virtual servers. You do not need to modify applications to use SMC to gain the performance benefits of communication by using SMC. Existing functions are preserved when SMC is used, such as the following functions:

- Load balancing, for example, sysplex distribution
- IP security zones
- Connection level security

Reference: See Shared Memory Communications reference information for a comprehensive list of materials for solutions that are related to Shared Memory Communications: SMC-R, SMC-D, and SMC Applicability Tool (SMC-AT).

Shared Memory Communications over Remote Direct Memory Access

Shared Memory Communications over Remote Direct Memory Access (RDMA), or Shared Memory Communications over RDMA (SMC-R), is a protocol solution that is based on sockets over RDMA. SMC-R enables TCP sockets applications to transparently use RDMA, which enables direct, high-speed, low-latency, memory-to-memory (peer-to-peer) communications. Communicating peers such as TCP/IP stacks dynamically learn about the shared memory capability by using traditional TCP/IP connection establishment flows, enabling the TCP/IP stacks to switch from TCP network flows to more optimized direct memory access flows that use RDMA.

RDMA is available on standard Ethernet-based networks by using the industry (InfiniBand Trade Association) standard referred to as RDMA over Converged Ethernet (RoCE). RoCE enables the use of both standard TCP/IP and RDMA solutions such as SMC-R over the same physical LAN fabric. SMC-R requires "RoCE Express" feature, which is sometimes referred to as an RDMA network interface card (RNIC). SMC-R provides an enterprise class of services for RDMA that are designed for IBM enterprise class data center networks.

As shown in Figure 60 on page 504, SMC-R enables two virtual servers that support RoCE to logically share memory through "RoCE Express" feature (shown as RNIC in Figure 60 on page 504) and over the RoCE network. When a virtual server that supports RoCE detects that a remote TCP connection partner supports shared memory communications, the connection is transparently and dynamically switched to use SMC-R protocols. The applications are unaware of the use of shared memory for communications.

© Copyright IBM Corp. 2000, 2022

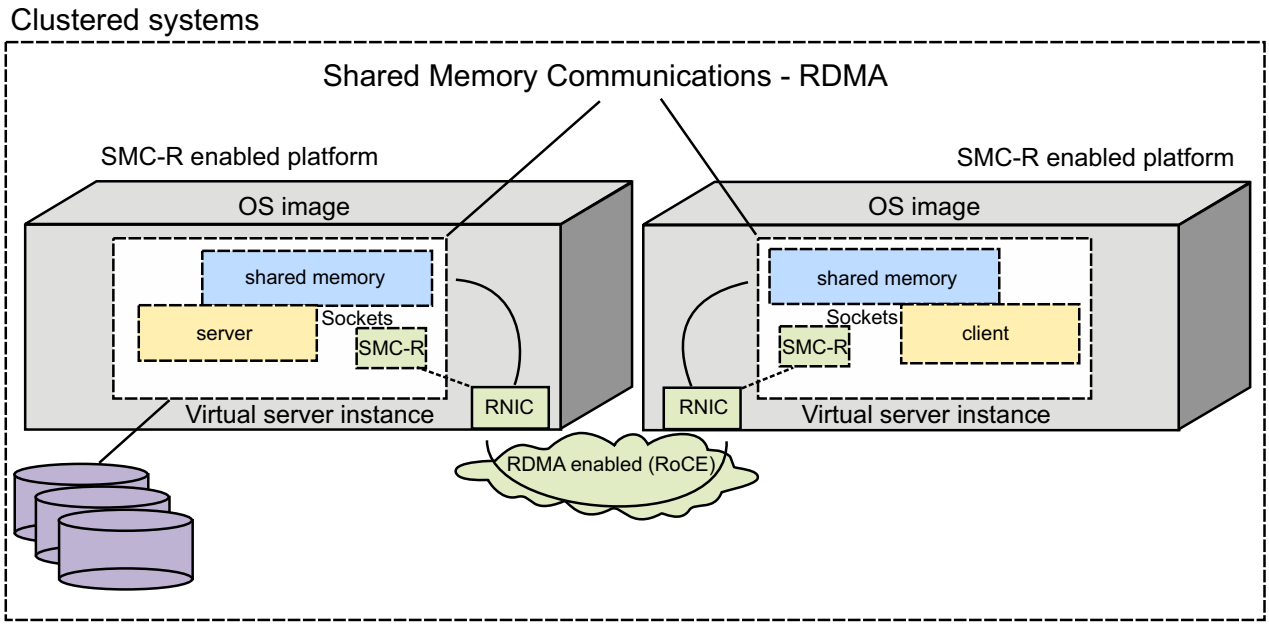


Figure 60. Shared Memory Communications over RDMA (SMC-R)

When redundant "RoCE Express" interfaces are provisioned, SMC-R also transparently provides the capability for failover processing when a failure is detected with SMC-R communications.

Remote Direct Memory Access over Converged Ethernet

Remote Direct Memory Access (RDMA) enables a host to make a subset of its memory directly available to a remote host. After RDMA connectivity is established between two TCP/IP stacks, either host can write to the memory of the remote host with no involvement from the remote host processors or operating system. RDMA enables efficient communications between the hosts because all the low-level functions are managed by RDMA network interface cards (RNICs) that are connected to each host, rather than by the software stack as is normally done for TCP/IP communications.

RDMA was traditionally confined to high-performance computing (HPC) environments where the cost of maintaining RDMA-capable network fabrics such as InfiniBand was justified given the emphasis of performance over cost. Now that RDMA is available on Ethernet fabrics through standards such as RDMA over Converged Ethernet (RoCE), the cost of adopting RDMA is lower because it can be enabled on the existing Ethernet fabrics that are used for IP network communications. Standard Ethernet management techniques are used to configure the RNIC adapters.

z/OS Communications Server provides support for sockets over RDMA by using SMC-R protocols. VTAM device drivers use Peripheral Component Interconnect Express (PCIe) operations to manage "RoCE Express" features that are defined to z/OS. Up to 16 "RoCE Express" PFID values can be defined to a z/OS TCP/IP stack.

Comparing 10 GbE RoCE Express feature environments

An IBM 10 GbE RoCE Express feature operates in either a dedicated or a shared RoCE environment.

z/OS Communications Server dynamically determines the operating environment supported by this generation of System z when the first 10 GbE RoCE Express feature is activated. Any additional 10 GbE RoCE Express features that are activated operate in the same environment that is determined when the first feature is activated.

Dedicated RoCE environment

In a dedicated RoCE environment, z/OS Communications Server uses PCIe Physical Function services to manage the 10 GbE RoCE Express feature.

A 10 GbE RoCE Express feature operating in a dedicated RoCE environment can only be used by a single LPAR. z/OS allows the feature to be shared by up to eight TCP/IP stacks within a single LPAR. Each TCP/IP stack uses the same PCIe function ID (PFID) value to define its representation of the 10 GbE RoCE Express feature. The PFID value is defined by using traditional hardware configuration definition (HCD) tools. In a dedicated RoCE environment, only one of the two available ports can be used at a time.

Figure 61 on page 505 is an example of a 10 GbE RoCE Express feature that is defined in a dedicated RoCE environment. A single z/OS image (z/OS 2) is using the 10 GbE RoCE Express features identified by PCHID values 100 and 200. Two PFID values (0001 and 0016) are defined to represent the features, and the PFID values correspond to the FID values defined in the HCD for the features. No other z/OS images can use these two features, although up to eight stacks on z/OS 2 can use the features. In this example, port 1 is used on both features, and port 2 cannot be used when port 1 is in use.

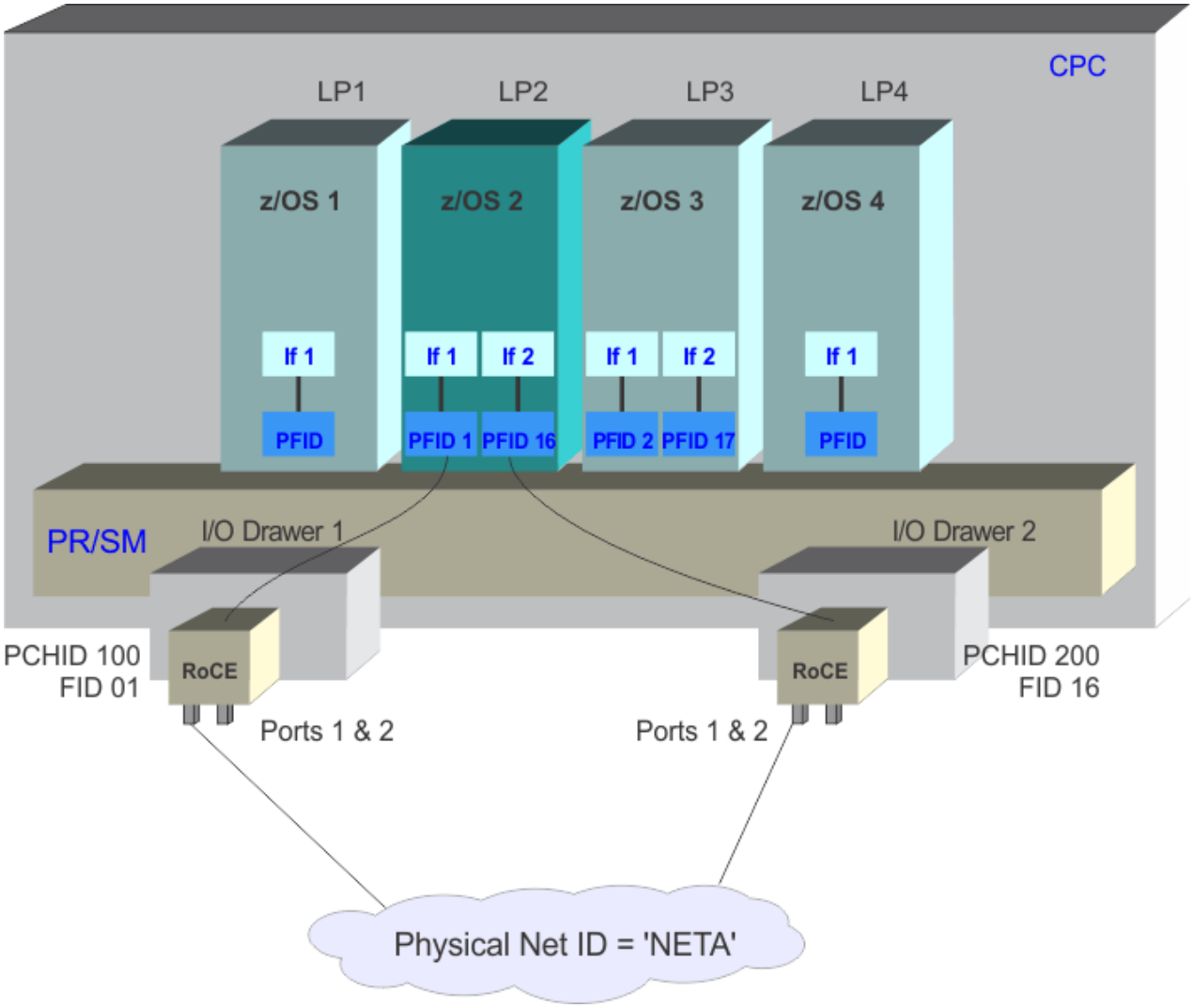


Figure 61. 10 GbE RoCE Express feature in a dedicated RoCE environment

A dedicated RoCE environment mode is supported on an IBM zEnterprise EC12 (zEC12) with driver 15, or an IBM zEnterprise BC12 (zBC12).

Shared RoCE environment

In a shared RoCE environment, z/OS Communications Server uses PCIe Virtual Function (VF) services to manage the 10 GbE RoCE Express feature, and System z provides the Physical Function management.

A 10 GbE RoCE Express feature operating in a shared RoCE environment can be shared by up to 31 operating system instances or TCP/IP stacks across the same central processor complex (CPC). Each TCP/IP stack within an LPAR, or each operating system instance, uses a unique FID value to define its representation of the 10 GbE RoCE Express feature. These FID values are defined by using HCD tools. In the shared environment, both 10 GbE RoCE ports can be used at the same time.

Guideline: For a TCP/IP stack, the FID value is represented by a PFID value on the GLOBALCONFIG statement in the TCP/IP profile. In addition, the same or different TCP/IP stacks can share the two 10 GbE RoCE Express ports of an individual 10 GbE RoCE Express feature if different PFID values are configured for the individual ports.

Figure 62 on page 506 is an example of a 10 GbE RoCE Express feature operating in a shared RoCE environment. Two z/OS images are using the 10 GbE RoCE Express features identified by PCHID values 100 and 200. Four unique PFID values are defined, two per z/OS image, to represent the usage of the features. The PFID values correspond to the FID values defined for the features in the HCD. In this example, the combination of PFID and port is unique for all four interfaces, but TCP/IP stacks are sharing the same feature and port.

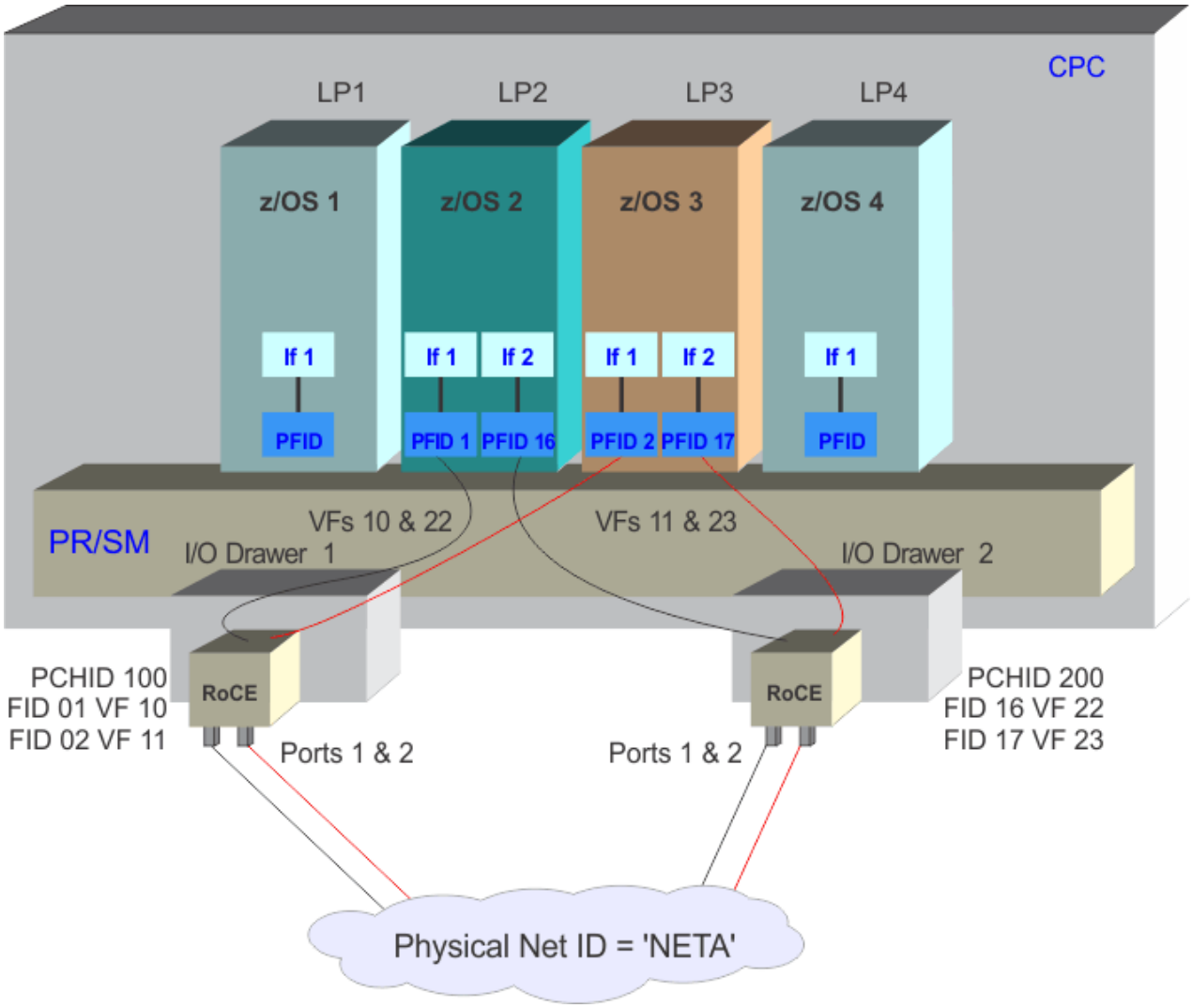


Figure 62. 10 GbE RoCE Express feature in a shared RoCE environment

Guideline: For full redundancy at a TCP/IP stack, configure PFID values that are associated with physically separate 10 GbE RoCE Express features. For example, in Figure 62 on page 506, for z/OS 2, the two features are in different System z I/O drawers. Therefore, the failure of one I/O drawer or one feature does not affect the other I/O drawer or feature.

A shared RoCE environment must be used on the IBM z13® (z13) or later systems.

RoCE Express2 feature environment

A RoCE Express2 feature operates in a shared RoCE environment. In general, the same rules and guidelines for defining and using a RoCE Express feature apply to a RoCE Express2 feature. Each RoCE Express2 port can be shared by 31 or 63 operating system instances or TCP/IP stacks across the same CPC, depending on the z14 model's capability. See Shared RoCE environment for a description of operating in a shared RoCE environment.

Shared Memory Communications - Direct Memory Access

Shared Memory Communications - Direct Memory Access (SMC-D) uses internal shared memory (ISM) for communication between two SMC capable peers that are located on the same central processor complex (CPC). The communicating peers, such as TCP/IP stacks, dynamically detect the shared memory capability by using traditional TCP/IP connection establishment flows. The shared memory enables the TCP/IP stacks to switch from TCP network flows to more optimized direct memory access flows that use ISM.

As shown in Figure 63 on page 507, SMC-D enables two virtual servers that support ISM to logically share memory. When a virtual server that supports ISM detects that a TCP connection partner also supports shared memory communications by using ISM, the connection is transparently and dynamically switched to use SMC-D protocols. The applications are unaware of the use of shared memory for communications.

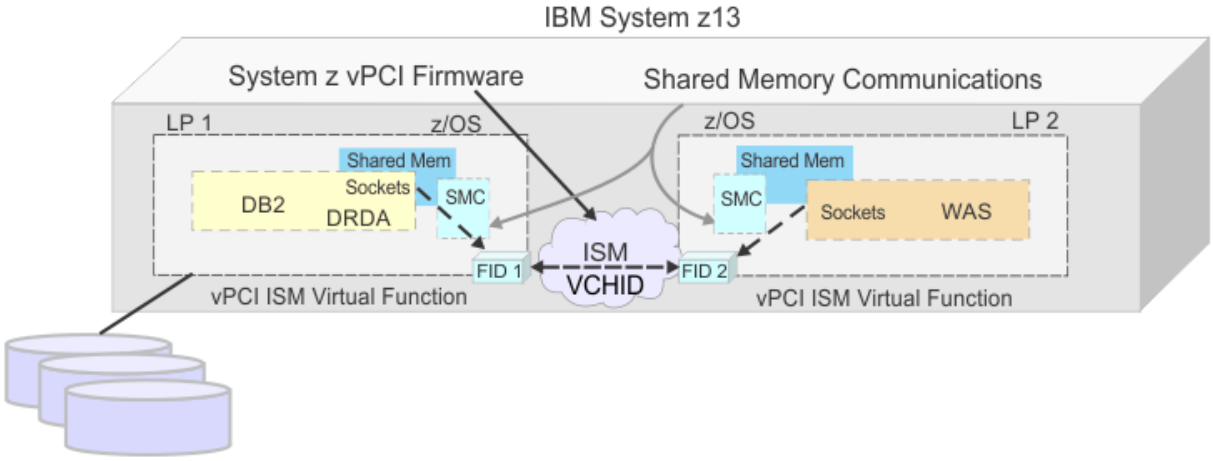


Figure 63. Shared Memory Communications - Direct Memory Access (SMC-D)

Shared Memory Communications concepts

The following concepts apply to Shared Memory Communications (SMC).

Rendezvous processing

Shared Memory Communications (SMC) is enabled by using the GLOBALCONFIG statement in the TCP/IP profile data set.

- Shared Memory Communications over RDMA (SMC-R) is enabled by specifying one or more Peripheral Component Interconnect Express (PCIe) function ID (PFID) values on the SMCR parameter of the

GLOBALCONFIG statement in the TCP/IP profile data set. Each PFID value represents a "RoCE Express" feature that is configured by using the traditional hardware configuration definition (HCD) tools. TCP/IP activates "RoCE Express" interfaces when the first SMC-R capable interface is started. SMC-R capable interfaces include IPAQENET or IPAQENET6 interfaces with the OSD channel path ID type. Any TCP connections that are routed over SMC-R capable interfaces are eligible for SMC-R communications.

- Shared Memory Communications - Direct Memory Access (SMC-D) is enabled by specifying the SMCD parameter of the GLOBALCONFIG statement in the TCP/IP profile data set. When SMC-D capable interfaces are activated, z/OS Communications Server selects an available ISM device that is associated with the same physical network as the SMC-D capable interface. Only one ISM device is activated for each physical network. The ISM device is represented by a Peripheral Component Interconnect Express (PCIe) function ID (PFID) value that is configured by using the traditional hardware configuration definition (HCD) tools. SMC-D capable interfaces include IPAQENET or IPAQENET6 interfaces with the OSD channel path ID type, and IPAQIDIO and IPAQIDIO6 HiperSockets interfaces. Any TCP connections that are routed over SMC-D capable interfaces are eligible for SMC-D communications.

The decision about whether an eligible connection uses SMC communications is reached during traditional TCP connection establishment. *Rendezvous processing* is the term that is used to describe the sequence of connection management flows (shown in Figure 64 on page 508) that are required to establish SMC-R communications between two peers.

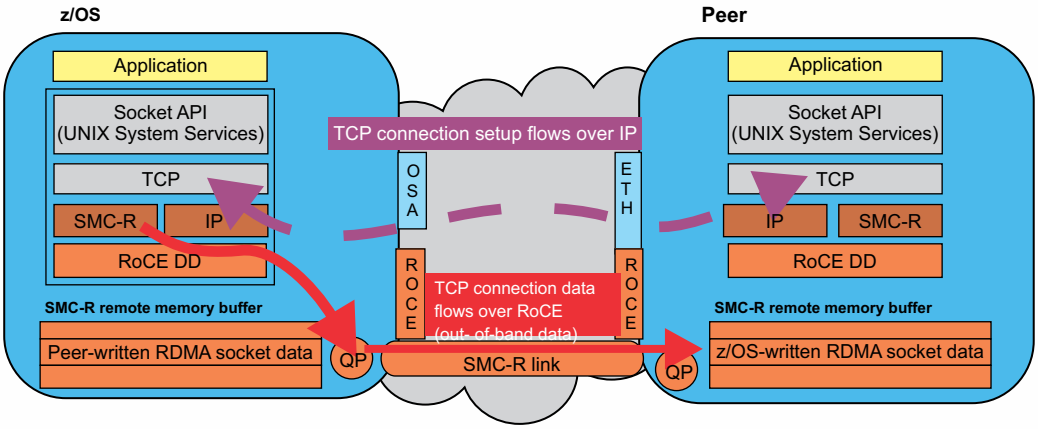


Figure 64. Rendezvous processing

The exchange of information occurs in different stages:

- Extra information in TCP connection establishment flows

  Applications still use the standard three-way handshake mechanisms to establish TCP connections. When SMC-R or SMC-D communications are enabled, the client and the server can indicate support for either SMC-R or SMC-D protocols:

  - The client adds TCP option 254 in the SYN request to indicate that SMC is supported by the client.
  - The server responds with TCP option 254 in the SYN-ACK to indicate that the server also supports SMC.

  No additional exchange of information is required in this stage of the rendezvous processing.

  Tip:

  Some firewalls can be configured to drop TCP options. The SMC protocol requires TCP option 254 to flow on the TCP/IP 3-way handshake. If you are enabling SMCv2 (SMC multiple IP subnet support) it is recommended that you verify that your firewalls permit TCP option 254 to flow. For more information about the SMC protocol, see RFC 7609.

  Tip:

  If you are enabling SMC-Rv2 with RoCEv2, it is also recommended that you verify that UDP port 4791 (RoCEv2) is open in your firewalls.

- In-band SMC Connection Layer Control (CLC) messages

  Conceptually, CLC messages are similar to the SSL handshake processing that occurs after the TCP connection is established.

  SMC-D protocols are preferred over SMC-R protocols. If both the client and the server indicate support for SMC-D processing, after the TCP connection is established, the client and the server negotiate the use of SMC-D for this TCP connection by using SMC-D CLC messages that flow as in-band data over the TCP connection.

  The SMC-D CLC messages exchange the following information:

  – Peer identification information
  – SMC-D buffer information that is necessary to exchange data between the peers

  For detailed description about the information that SMC-D CLC messages exchange, see "SMC-D links" on page 513.

  If the attempt to establish SMC-D communications succeeds, the SMC-D link is established, and the peers can start exchanging data by using the SMC-D buffers.

  If SMC-D communications is not available, and both the client and the server indicate support for SMC-R processing, the client and the server negotiate the use of SMC-R for this TCP connection by using SMC-R CLC messages that flow as in-band data over the TCP connection.

  The SMC-R CLC messages exchange the following information:

  – Network routing information
  – RDMA over Converged Ethernet (RoCE) routing credentials
  – SMC-R buffer information that is necessary to select or establish the RoCE path between the peers

  For detailed description about the information that SMC-R CLC messages exchange, see "SMC-R links" on page 510.
- SMC-R Link Layer Control (LLC) messages

  After the SMC-R information is exchanged, SMC-R LLC messages are exchanged across the RoCE fabric to confirm that the RoCE information is correct and that the remote memory can be accessed. This stage is skipped if an existing RoCE connection is used for this TCP connection.

The TCP/IP stack does not allow the client and server applications to exchange application data during rendezvous processing. Because no application data is exchanged, the TCP connection can revert to IP protocols if there is a failure during the setup of the SMC communications. However, after the TCP connection is committed to using SMC protocols, the TCP connection cannot fall back to using IP protocols if SMC communications encounter an error.

- For SMC-R communications, the TCP connection is committed to use SMC logic after the RoCE connection is confirmed by using the SMC-R LLC messages.
- For SMC-D communications, the TCP connection is committed to use SMC logic after the successful exchange of SMC-D CLC messages.

Both the client and the server nodes maintain a series of rendezvous timers to ensure that the rendezvous processing completes in time. If one of the timed events does not complete as expected, the TCP connection reverts to using IP protocols. However, future TCP connections can still attempt to use SMC.

To reduce the overhead of persistent rendezvous failures (TCP connections reverting to using IP protocols) to the same destination IP address, you can use the default AUTOCACHE function. This function is controlled by the AUTOCACHE and NOAUTOCACHE subparameters of the SMCGLOBAL parameter on the GLOBALCONFIG profile statement and is set to AUTOCACHE by default. The AUTOCACHE function caches rendezvous failures per IP address destination. If the AUTOCACHE function detects too many rendezvous failures to a specific IP address, the function prevents additional rendezvous attempts to that IP address. The AUTOCACHE function is started only when you enable SMC. For more information about enabling SMC, see the description of the GLOBALCONFIG SMCR and SMCD parameters in z/OS Communications Server: IP Configuration Reference.

Even though the data is sent out of band with SMC communications, the TCP connection remains active, primarily to facilitate connection termination processing. If you have TCP server applications that primarily use many short-lived TCP connections, you might want to avoid rendezvous processing. You can prevent these server applications from using SMC in one of the following ways:

Use the AUTOSMC monitoring function
The AUTOSMC monitoring function dynamically monitors incoming TCP connections to local TCP server applications and determines whether the use of SMC is beneficial for the workload. This function is configured by the AUTOSMC and NOAUTOSMC subparameters of the SMCGLOBAL parameter on the GLOBALCONFIG profile statement and is set to AUTOSMC by default. However, the AUTOSMC monitoring function is started only when you enable SMC. For more information about enabling SMC, see the description of the GLOBALCONFIG SMCR and SMCD parameters in z/OS Communications Server: IP Configuration Reference. If the function determines that most of the incoming connections are short-lived and exchange a small amount of data then SMC processing will be bypassed for all new connections to this server. This monitoring is dynamic in nature so that it can detect changes in workload patterns. You can monitor the results of this dynamic monitoring and SMC enablement/disablement using the Netstat ALL/-A command. For more information about the Netstat ALL/-A command, see z/OS Communications Server: IP System Administrator's Commands.

Disable SMC eligiblity for the port
Explicitly specify the NOSMC parameter on the PORT or PORTRANGE statements that define the port or ports that the server application uses. For more information, see z/OS Communications Server: IP Configuration Reference.

Shared Memory Communications links and link groups

After two Shared Memory Communications (SMC) peers recognize during rendezvous processing that shared memory communications are possible, the peers create SMC links and, in the case of SMC-R, SMC link groups.

SMC-R links

After two Shared Memory Communications over RDMA (SMC-R) peers recognize during rendezvous processing that shared memory communications are possible, a logical point-to-point SMC-R link is established between the stacks over the RDMA over Converged Ethernet (RoCE) fabric. An SMC-R link, as shown in Figure 65 on page 511, is uniquely defined by a combination of the following information:

- Remote and local virtual MAC (VMAC) values

  A VMAC is a 6-byte value that is a virtual representation of the physical MAC address for a "RoCE Express" interface (shown as RNIC in Figure 65 on page 511).

  Each TCP/IP stack that activates a particular "RoCE Express" interface is assigned a different VMAC value.

- Remote and local global ID (GID) values

  A GID is a 16-byte value. z/OS Communications Server generates the GID values by converting the VMAC address of the "RoCE Express" interface into an IPv6 link-local address.

- Remote and local queue pair (QP) values

  A QP represents one end of the logical connection between two RDMA peers. A combination of two reliable connected queue pairs (RC QPs) forms a single logical point-to-point link. The link enables exactly one pair of communicating RDMA peers to send and receive messages and initiate RDMA activities between themselves. A "RoCE Express" interface associates units of work, such as confirmation of sent data or indication of received data, to a specific QP to enable the SMC-R protocols to identify which TCP/IP stack to notify for the unit of work. The stack then determines which TCP connection that uses that RC QP is to process the data.

- Virtual LAN (VLAN) ID

  You can optionally use VLANs to isolate application traffic into different virtual networks on the same physical Ethernet.

- If you use VLANs, the VLAN ID specified on the IPAQENET or IPAQENET6 INTERFACE statement is used as an attribute to create unique SMC-R links between the peers for unique VLANs. In other words, the SMC-R links are VLAN qualified.
- If you do not use VLANs, no VLAN ID is used to define the SMC-R links between the peers. In other words, the SMC-R links are not VLAN qualified.

For more information about using VLANs, see "SMC-R VLANID usage for the 10 GbE RoCE Express feature" on page 519.

Application traffic between the two peers that uses the same remote and local VMACs, GIDs, and QPs, and that is associated with the same VLAN when VLANs are defined, can use the same SMC-R link.

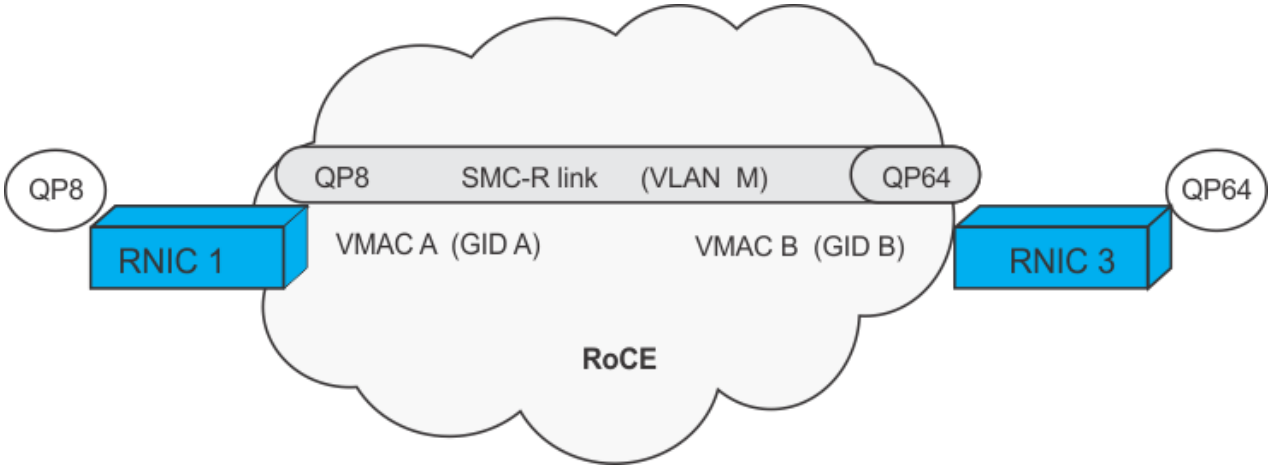


Figure 65. Identifying an SMC-R link

In addition to the 7-tuple (local VMAC, GID, QP# + remote VMAC, GID, QP# + VLAN ID) that uniquely defines an SMC-R link, each peer assigns a 4-byte SMC-R link ID value that uniquely identifies the SMC-R link within its own resource space. This SMC-R link ID is exchanged between peers and is intended to be used for network management and diagnostic purposes. For instance, you can use the SMC-R link ID to filter Netstat report information that is related to a specific SMC-R link. For more information, see "Displaying SMC information" on page 560.

An SMC-R link supports multiple TCP connections between the same two peers, as shown in Figure 66 on page 511. The first TCP connection between the peers establishes the SMC-R link, and subsequent TCP connections between the peers can use the previously established link. Because subsequent TCP connections between the peers can use the previously established link, extra SMC-R link setup costs between the peers are avoided.

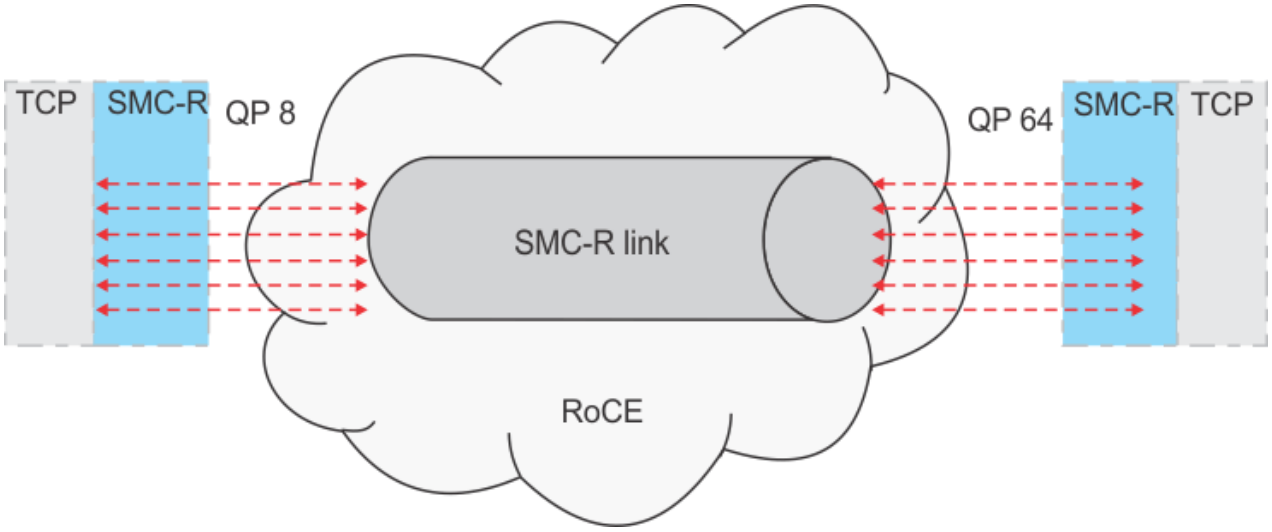


Figure 66. Multiple TCP connections over one SMC-R link

SMC-R link groups

A Shared Memory Communications over RDMA (SMC-R) link group is a logical grouping of SMC-R links between two communicating peers, as shown in Figure 67 on page 512. An SMC-R link group is formed when the initial SMC-R link is established between two peers.

All SMC-R links in an SMC-R link group must be *equal* links. SMC-R links are considered to be equal when all of the following conditions are true:

- The links provide access to the same RDMA memory buffers at the remote peer virtual servers.
- The links have the same VLAN ID, or they do not use a VLAN ID.
- The links have the same TCP server and TCP client roles or relationship.

  A peer that is acting as the TCP connection server has different responsibilities for establishing and maintaining SMC-R communications than a peer that is acting as the TCP connection client. Unique SMC-R link groups are established between two peers when the peers act as both servers and clients for TCP connections.

When the initial SMC-R link is established and a second "RoCE Express" interface is available, Communications Server establishes an equal SMC-R link between the peers. The "RoCE Express" interfaces are shown as RNICs in Figure 67 on page 512.

Adding a second SMC-R link to the SMC-R link group provides the following benefits:

- High availability

  To maintain high availability, you need two SMC-R links between SMC-R peers. If a failure occurs with one SMC-R link, TCP connections that are using the failing SMC-R link are switched to the other active link in the link group and disruptions to application workloads are avoided. For more information, see “SMC-R high availability considerations” on page 522.
- Workload balancing

  TCP connections are distributed across the SMC-R links in a link group, increasing bandwidth and avoiding bottlenecks.

  Rule: Workload balancing within an SMC-R link group occurs only when both the local and the remote peers have two "RoCE Express" interfaces, and thus two SMC-R links are established in the link group.

  Guideline: When you are provisioning PFIDs for a specific PNetID for each TCP/IP stack, you should assign "RoCE Express" PFIDs with the same link speed (i.e. 10 GbE, 25 GbE). If this guideline is not followed, this will result in SMC-R link groups being created with SMC-R links with unequal throughput capacities. During error recovery scenarios, fail over processing might overload the lower capacity SMC-R link that could result in recovery failures.

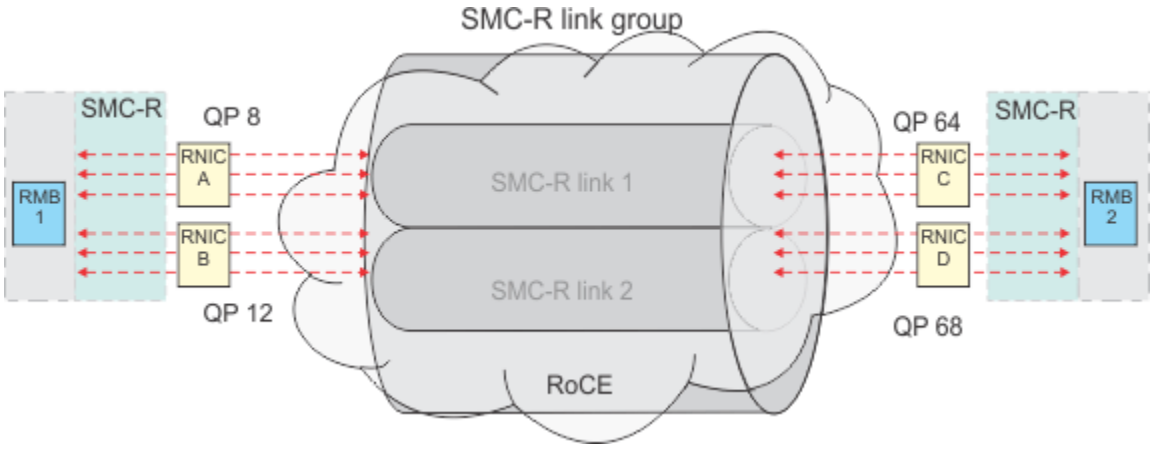


Figure 67. SMC-R link group

Because SMC-R links within a link group are considered equal, TCP connections can be assigned to any SMC-R link within the group. Furthermore, the client and the server can choose to assign the TCP

connection to different SMC-R links within the group, and can move the TCP connections from one SMC-R link to another within the group. For example, in Figure 67 on page 512, client traffic might flow over one SMC-R link (between RNICs A and C) and server traffic might flow over the other SMC-R link (between RNICs B and D). The peers do not have to exchange knowledge of which physical "RoCE Express" interface is being used for data transmission, and the recipient is only aware that data was placed into the RDMA memory buffer.

An SMC-R link group remains active for up to 10 minutes after the last TCP connection that is using the link group is stopped.

SMC-D links

After two Shared Memory Communications - Direct Memory Access (SMC-D) peers recognize during rendezvous processing that shared memory communications are possible, a logical SMC-D link is established between the peers by using internal shared memory (ISM). An SMC-D link, as shown in Figure 68 on page 513, is uniquely defined by a combination of the following information:

- Remote and local global ID (GID) values

  An SMC-D GID is an 8-byte value that the ISM device assigns.
- Virtual LAN (VLAN) ID

  You can optionally use VLANs to isolate application traffic into different virtual networks on the same physical Ethernet or HiperSockets CHPID.

  – If you use VLANs, the VLAN ID that is specified on the IPAQENET, IPAQENET6, IPAQIDIO, or IPAQIDIO6 INTERFACE statement is used as an attribute to create unique SMC-D links between the peers for unique VLANs. In other words, the SMC-D links are VLAN qualified.

  – If you do not use VLANs, no VLAN ID is used to define the SMC-D links between the peers. In other words, the SMC-D links are not VLAN qualified.

  For more information about using VLANs, see “VLANID considerations” on page 518.

If the application traffic that is between the two peers uses the same remote and local GIDs, and is associated with the same VLAN when VLANs are defined, the application traffic can use the same SMC-D link.

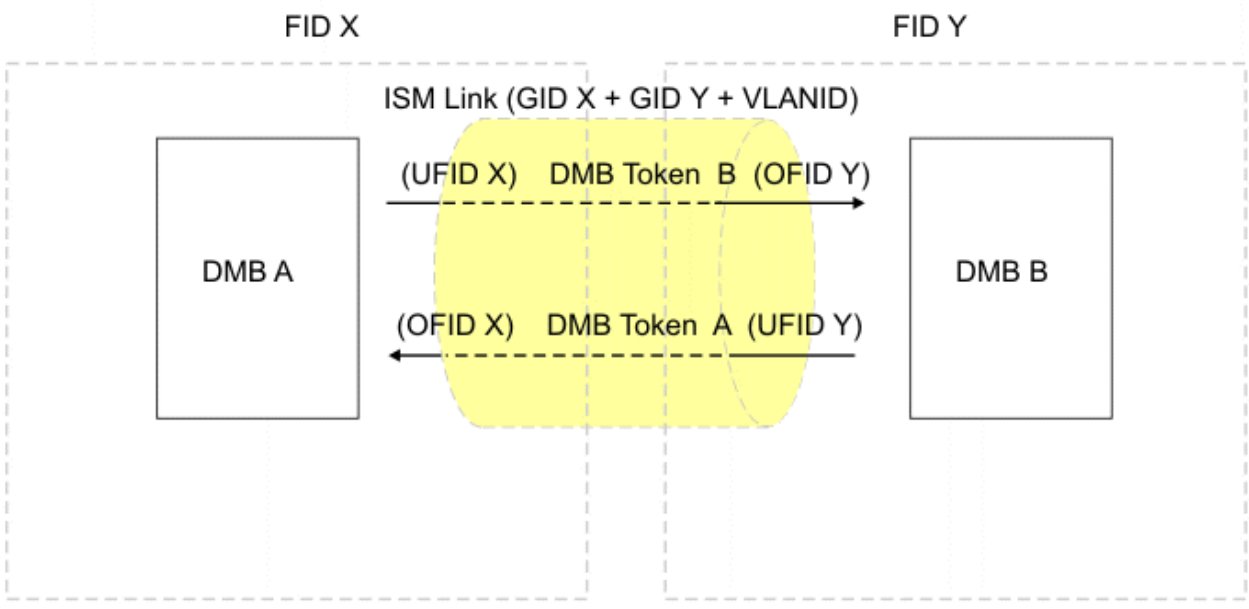


Figure 68. Identifying an SMC-D link

Besides the 3-tuple (local GID + remote GID + VLAN ID) that uniquely defines an SMC-D link, each peer assigns a 4-byte SMC-D link ID that uniquely identifies the SMC-D link within its own resource space. This

SMC-D link ID is exchanged between peers and can be used for network management and diagnostic purposes. For instance, you can use the SMC-D link ID to filter Netstat report information that is related to a specific SMC-D link. For more information, see “Displaying SMC information” on page 560.

An SMC-D link supports multiple TCP connections between the same two peers. The first TCP connection between the peers establishes the SMC-D link, and subsequent TCP connections between the peers can use the previously established link.

SMC memory buffers

Each stack locally allocates memory for an SMC memory buffer to receive inbound data that uses SMC communications. An SMC memory buffer that is used for SMC-R communications is called a remote memory buffer (RMB), and an SMC memory buffer that is used for SMC-D communications is called a direct memory buffer (DMB). The sending operating system places TCP socket application data directly into the memory buffer that the receiving stack assigns to receive data for a TCP connection. The receiving stack then copies the data from the memory buffer into the receive buffer of the receiving socket application.

A memory buffer is partitioned into elements of equal size, and each element is associated with a single TCP connection. Direct memory buffers are partitioned into direct memory buffer elements (DMBEs), and remote memory buffers are partitioned into remote memory buffer elements (RMBEs). In the Connection Layer Control (CLC) messages during rendezvous processing, each peer communicates the location of its local DMBE, or its RMBE and a remote key, to allow remote access to the memory buffer. The remote host has write access and the local host reads the data for passing to the socket application.

SMC-R usage of memory buffers

For z/OS Communications Server, an RMB is a contiguous, 1-MB block of fixed 64-bit private storage. In Figure 69 on page 514, an RMB is created on z/OS, into which the peer node can write, and an RMB is created on the peer node, into which z/OS can write. An RMB must be registered with the "RoCE Express" interface so that the storage is available to the remote peer.

Each RMB is associated with the reliable connected queue pairs (RC QPs) and therefore with the SMC-R link for the two communicating peers. The association of the RMB to the RC QPs ensures that only the two peers have RDMA access to this particular RMB. In addition, all RMBs that are associated with a particular SMC-R link must be accessible to the remote peer by using any SMC-R link in the link group. This accessibility enables the remote peer to use any link within the link group to place the TCP connection data into the correct RMB.

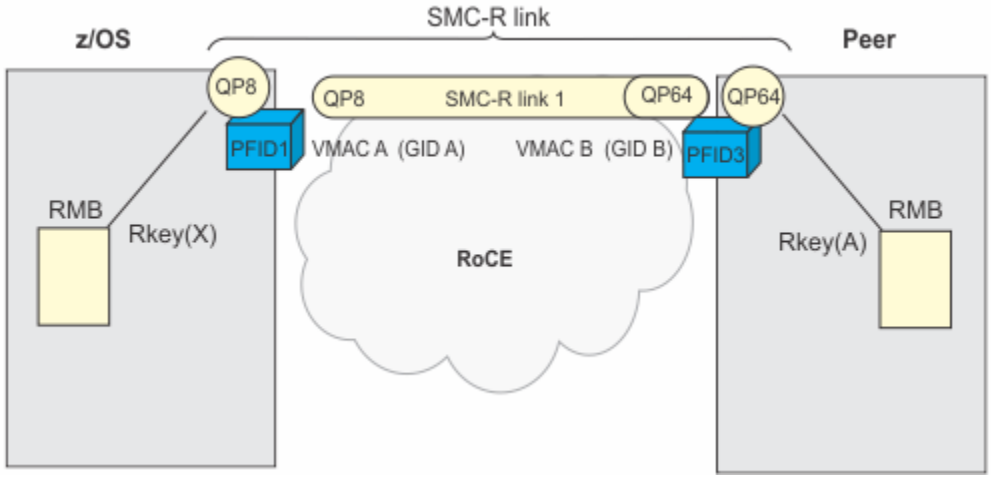


Figure 69. RMBs assigned to an SMC-R link

The SMC-R peers are not required to assign RMBEs of the same size for the client and the server of a specific TCP connection. The peers can use one or more RMBs for the TCP connections that are using the same SMC-R link.

When an SMC-R link group is first established, z/OS Communications Server allocates a basic set of three RMBs for the link group. The RMBE sizes for the RMBs are not defined initially, but instead are determined by the needs of the TCP connections that use the SMC-R link. z/OS Communications Server uses RMBE sizes of 32 KB, 64 KB, 128 KB, 256 KB, and greater than 256 KB. The appropriate size for the TCP connection is selected based on the receive buffer size of the local application. When the application does not explicitly set the buffer size by using SETSOCKOPT(), the default value is determined by the value of the TCPRCVBUFRSIZE parameter on the TCPCONFIG statement.

More RMBs are allocated as needed to accommodate more TCP connections or different receive buffer sizes.

SMC-D usage of memory buffers

For z/OS Communications Server, a DMB is a contiguous, 1 MB block of pinned 64-bit private storage. A DMB is created on z/OS, into which the peer node can write, and a DMB is created on the peer node, into which z/OS can write. A DMB must be registered with the ISM interface so that the storage is available to the remote peer. Each DMB is associated with one and only one SMC-D link.

The SMC-D peers are not required to assign DMBEs of the same size for the client and the server of a specific TCP connection. The peers can use one or more DMBs for the TCP connections that are using the same SMC-D link.

When an SMC-D link is first established, z/OS Communications Server allocates a basic set of three DMBs for the link. The DMBE sizes for the DMBs are not defined initially, but instead are determined by the needs of the TCP connections that use the SMC-D link. z/OS Communications Server uses DMBE sizes of 64 KB, 128 KB, 256 KB, and greater than 256 KB. The appropriate size for the TCP connection is selected based on the receive buffer size of the local application. When the application does not explicitly set the buffer size by using SETSOCKOPT(), the default value is determined by the value of the TCPRCVBUFRSIZE parameter on the TCPCONFIG statement.

More DMBs are allocated as needed to accommodate more TCP connections or different receive buffer sizes.

SMC-R staging buffers

Each TCP/IP stack allocates 64-bit fixed private staging buffer memory for outbound Shared Memory Communications over RDMA (SMC-R) data. Staging buffers are not associated with specific SMC-R links or link groups, but are used by all TCP connections that traverse SMC-R links on this stack.

Like remote memory buffers (RMBs), staging buffers are 1 MB in length and must be registered with the "RoCE Express" interface before they are used for SMC-R communications. Unlike RMBs, the address of the staging buffer is not shared with the remote SMC-R peer because only local applications access the staging buffer storage.

z/OS Communications Server allocates 4 MB of staging buffer storage when the first "RoCE Express" interface is activated, and allocates more buffers as necessary to accommodate application workloads.

SMC filters

Shared Memory Communications filters are provided to allow the administrator to have a more granular level of control of which TCP connections are eligible to use SMC. SMC filters control all forms of SMC. For more details, see “SMC filters” on page 539.

Shared Memory Communications terms

The following terms apply to Shared Memory Communications (SMC). You can use this list as needed for brief descriptions when you are using other SMC information.

Associated ISM interface
An internal shared memory (ISM) interface that is associated with an Shared Memory Communications - Direct Memory Access (SMC-D) capable interface that has the same physical network ID.

Associated RNIC interface
An IBM 10 GbE RoCE Express interface that is associated with an Shared Memory Communications over Remote Direct Memory (SMC-R) capable interface that has the same physical network ID.

Direct memory buffer (DMB)
Local memory that is used to receive inbound data over an SMC-D link. The remote peer places TCP socket application data directly into the DMB that the local peer assigns to receive data for the TCP connection. The local peer then copies the data from the DMB into the receive buffer of the receiving socket application.

DMB element (DMBE)
The portion of a DMB that is associated with a specific TCP connection. Each DMB is partitioned into one or more DMBEs.

IBM 10 GbE RoCE Express feature or RoCE Express2 feature
A feature that enables Remote Direct Memory Access by managing low-level functions that the TCP/IP stack typically handles.

IBM 10 GbE RoCE Express interface
An interface that is dynamically created by TCP/IP that uses a particular port of an IBM 10 GbE RoCE Express feature.

IBM RoCE Express2 interface
An interface that is dynamically created by TCP/IP that uses a particular port of a 10 GbE or 25 GbE RoCE Express2 feature.

Internal path
The System z internal PCIe infrastructure for "RoCE Express" features. The internal path of a "RoCE Express" feature is determined based on how the feature is plugged into the System z I/O drawers.

ISM device
The System z firmware that facilitates the use of internal shared memory for SMC-D processing. An ISM device is identified by a unique PCIe function ID (PFID) and a virtual channel ID (VCHID). An ISM device is configured in the hardware configuration definition (HCD) and is associated with a single physical network ID.

ISM interface
An interface that is dynamically created by TCP/IP to represent an ISM device for SMC-D capable interfaces with the same physical network ID.

Operating system images
Logical partitions (LPARs) or guest virtual machines that operate in the same central processor complex (CPC).

Physical channel ID (PCHID)
A 2-byte hexadecimal value that is used to uniquely define a RoCE Express feature.

PCIe function ID (PFID)

A value that represents the SMC device.

For SMC-R, the PFID value is configured on the SMCR parameter of the GLOBALCONFIG statement in the TCP/IP profile to identify a "RoCE Express" feature. The PFID represents a physical "RoCE Express" feature and must match a FID value configured in the hardware configuration definition (HCD) for the PCHID value that identifies the feature. When the "RoCE Express" feature is installed on a System z that supports a shared RoCE environment, the same physical feature can be shared with other operating system images, and multiple PFID values specified on the same GLOBALCONFIG statement can represent different ports on the same physical "RoCE Express" feature.

For SMC-D, the PFID value is obtained during ISM interface activation that identifies an ISM device. The PFID value matches an FID value that is configured in the HCD for the VCHID value that represents the ISM device.

Peripheral Component Interconnect Express (PCI Express, or PCIe)
A local bus that provides the high-speed data path between the processor and an SMC device. For SMC-R, the device is an attached "RoCE Express" feature. For SMC-D, the device is the ISM device.

Physical network ID (PNetID)
A value that is defined to uniquely identify your physical layer 2 LAN fabric or physical broadcast domain. You can use this value to logically associate the System z features, adapters, and ports to be physically connected to your network. You specify the PNetID in a single step within the hardware configuration definition (HCD), and all operating systems of all associated central processor complexes (CPCs) can dynamically learn and use this definition.

RDMA network interface card (RNIC)
A RoCE Express feature that enables Remote Direct Memory Access by managing low-level functions that are typically handled by the TCP/IP stack.

RDMA over Converged Ethernet (RoCE)
An InfiniBand Trade Association (IBTA) standard that enables Remote Direct Memory Access over Converged Ethernet.

Redundancy level
For an SMC-R link group, this value indicates the level to which z/OS Communications Server can provide dynamic failover processing if there is a failure of an underlying "RoCE Express" interface or the associated network hardware.

Reliable connected queue pair (RC QP)
A logical connection between two virtual servers that enables that specific pair of servers to use RDMA communications between themselves.

Remote Direct Memory Access (RDMA)
A high-speed, low-latency network communications protocol in which data is transferred directly to the memory of a remote host with no involvement from the remote host processors or operating system.

Remote memory buffer (RMB)
Local memory that is used to receive inbound data over an SMC-R link. The remote peer places TCP socket application data directly into the RMB that the local peer assigns to receive data for the TCP connection. The local peer then copies the data from the RMB into the receive buffer of the receiving socket application.

Rendezvous processing
The sequence of TCP connection management flows that are required to establish SMC communications between two peers.

RMB element (RMBE)
The specific portion of an RMB that is associated with a specific TCP connection. Each RMB is partitioned into RMBEs.

RoCE environments
Depending on the level of hardware that is used, the 10 GbE RoCE Express feature operates in either a shared or a dedicated RoCE environment. A RoCE Express2 feature always operates in a shared RoCE environment.

Dedicated RoCE environment
A dedicated RoCE environment applies to an IBM zEnterprise EC12 (zEC12) with driver 15, or an IBM zEnterprise BC12 (zBC12). In this environment, only a single operating system instance can use a physical 10 GbE RoCE Express feature. Multiple operating system instances cannot concurrently share the feature.

Shared RoCE environment
A shared RoCE environment applies to an IBM z13 (z13) or later system. In this environment, multiple operating system instances can concurrently use or share the same physical RoCE Express feature. With IBM z13 (z13) or later systems, the RoCE Express feature operates in a shared environment even if only one operating system instance is configured to use the feature.

RoCE Express
Generic term for either IBM 10 GbE RoCE Express, IBM 10 GbE RoCE Express2, or IBM 25 GbE RoCE Express2 feature.

RoCE Express2
Generic term for either IBM 10 GbE RoCE Express2 or IBM 25 GbE RoCE Express2 feature.

SMC-D link
A logical representation of communication between two virtual servers that use SMC-D protocols. The concept of an SMC-D link exists primarily for operational consistency with SMC-R processing.

SMC-R link
A logical point-to-point link between two virtual servers that is used for SMC-R communications.

SMC-R link group
A logical grouping of equal SMC-R links between two communicating peers.

Staging buffer
Memory that the TCP/IP stack allocates for outbound SMC-R data. Staging buffers are not associated with specific SMC-R links or link groups, and are used by all TCP connections that traverse SMC-R links on this stack. Only local applications access the staging buffer storage.

Virtual channel ID (VCHID)
A 2-byte hexadecimal value that is used to uniquely define an ISM device.

Using Shared Memory Communications

There are some configuration considerations and environment setup steps before you can configure and use Shared Memory Communications over RDMA (SMC-R) or Shared Memory Communications - Direct Memory Access (SMC-D).

Tip: You can use the SMC Applicability Tool (SMCAT) report to better understand the amount of your TCP workload that can use SMC communications. For more information about the SMCAT, see z/OS Communications Server: IP System Administrator's Commands.

Configuration considerations for Shared Memory Communications

Before you configure Shared Memory Communications, consider the following factors:

- Decide whether to use VLANs. For more information, see “VLANID considerations” on page 518.
- Identify the physical connections between stacks and SMC devices. For more information, see “Physical network considerations” on page 520.
- Provide physical redundancy for high availability when using SMC-R. For more information, see “SMC-R high availability considerations” on page 522.
- Verify the system and network requirements.
  - For more information about system requirements for SMC-R, see “System requirements for SMC-R in a dedicated RoCE environment” on page 527, “System requirements for SMC-R in a shared RoCE environment” on page 527, and “Network requirements for SMC-R” on page 528.
  - For more information about system requirements for SMC-D, see “System requirements for SMC-D” on page 529.

VLANID considerations

The VLANID operand is optional on the following SMC capable interfaces:

- IPAQENET and IPAQENET6 INTERFACE statements with the OSD channel path ID type (CHPIDTYPE OSD)
- IPAQIDIO and IPAQIDIO6 INTERFACE statements (SMC-D support only)

On a specific OSA or HiperSockets transport resource list element (TRLE) basis, z/OS Communications Server enforces consistent VLAN definitions for INTERFACE statements that are associated with the same OSA or HiperSockets TRLE.

For example, when VLANs are not used, the stack configuration allows only a single INTERFACE statement, and the VLANID operand is omitted on that INTERFACE statement. When VLANs are used, multiple INTERFACE statements are allowed and each INTERFACE statement must specify a unique VLANID value.

The OSD VLAN attributes of the IPAQENET or IPAQENET6 interface are propagated to the "RoCE Express" interfaces (associated RNIC) or ISM interfaces (associated ISM) that have the same physical network identifier (PNetID) value. The HiperSockets VLAN attributes on the IPAQIDIO or IPAQIDIO6 interface are propagated to the ISM interfaces that have the same PNetID value. See Physical network considerations for more details on PNetID.

SMC-R VLANID usage for the 10 GbE RoCE Express feature

Whether SMC-R communications use virtual LANs depends on the definition of the SMC-R capable OSD interfaces that are extended to the associated 10 GbE RoCE Express interfaces. The 10 GbE RoCE Express feature can be shared by TCP/IP stacks that are configured to use different VLAN capabilities for the 10 GbE RoCE Express feature. You can use up to 126 unique VLANID values per 10 GbE RoCE Express port.

Depending on the operating mode, the number of VLANID values that can be used per 10 GbE RoCE Express feature has the following limits:

- When the 10 GbE RoCE Express feature operates in a dedicated RoCE environment, up to 126 unique VLANID values can be used per port.
- When the 10 GbE RoCE Express feature operates in a shared RoCE environment, up to 126 unique VLANID values can be used per port. In addition, each virtual function (VF) PFID representation of the feature can use up to 16 VLANID values, although internal 10 GbE RoCE Express feature limitations might further reduce that maximum value for individual PFID representations.

  Result: Multiple VF representations of the same 10 GbE RoCE Express feature can use the same VLANID value, and only one of the available 126 VLANID values is used.

Result: If you define more unique VLANID values for one PNetID on the SMC-R capable INTERFACE statements than the 10 GbE RoCE Express feature can support, the VLANID values of the last INTERFACE statements to be activated are not registered with the 10 GbE RoCE Express feature. The IPAQENET or IPAQENET6 interfaces can start, but TCP connections that are established over these interfaces cannot use SMC-R communications. Netstat ALL/-A reports that display the TCP connections include the following diagnostic information for the connection:

```
SMCRSTATUS:        INACTIVE
SMCREASON:        00005206 - VLAN ID NOT FOUND
```

SMC-R VLANID usage for the RoCE Express2 feature

Whether SMC-R communications use virtual LANs depends on the definition of the SMC-R capable OSD interfaces that are extended to the associated RoCE Express2 interfaces. The RoCE Express2 feature can be shared by TCP/IP stacks that are configured to use different VLAN capabilities for the RoCE Express2 feature. There is no limit on the number of unique VLANIDs that you can use per RoCE Express2 port.

SMC-D VLANID usage

Whether SMC-D communications use virtual LANs depends on the definition of the SMC-D capable OSD or HiperSockets interfaces that are extended to the associated ISM interfaces. The ISM device can be shared by TCP/IP stacks that are configured to use different VLAN capabilities for the ISM device. Each TCP/IP stack, or virtual function (VF), can use up to 64 unique VLANID values per ISM interface.

Result: If you define more unique VLANID values for one PNetID on the SMC-D capable INTERFACE statements than the ISM device can support, the VLANID values of the last INTERFACE statements to be

activated are not registered with the ISM device. The IPAQENET, IPAQENET6, IPAQIDIO, or IPAQIDIO6 interfaces can start, but TCP connections that are established over these interfaces cannot use SMC-D communications. Netstat ALL/-A reports that display the TCP connections include the following diagnostic information for the connection:

```
SMCDSTATUS:        INACTIVE
SMCREASON:        00005206 - VLAN ID NOT FOUND
```

The usage of VLAN IDs for ISM have changed for SMC Version 2. For details of the SMC-Dv2 VLAN ID usage for ISM, see “SMC-Dv2 and ISM VCHID Association using PNetIDs” on page 541.

Physical network considerations

The TCP/IP stack must be able to determine which physical network is connected to a particular "RoCE Express" or ISM interface, so that the interface can be associated with the SMC capable IPAQENET, IPAQENET6, IPAQIDIO, and IPAQIDIO6 interfaces that connect to that same physical network. For example, in Figure 70 on page 520, three distinct and physically separated networks can be accessed by using SMC-R communications.

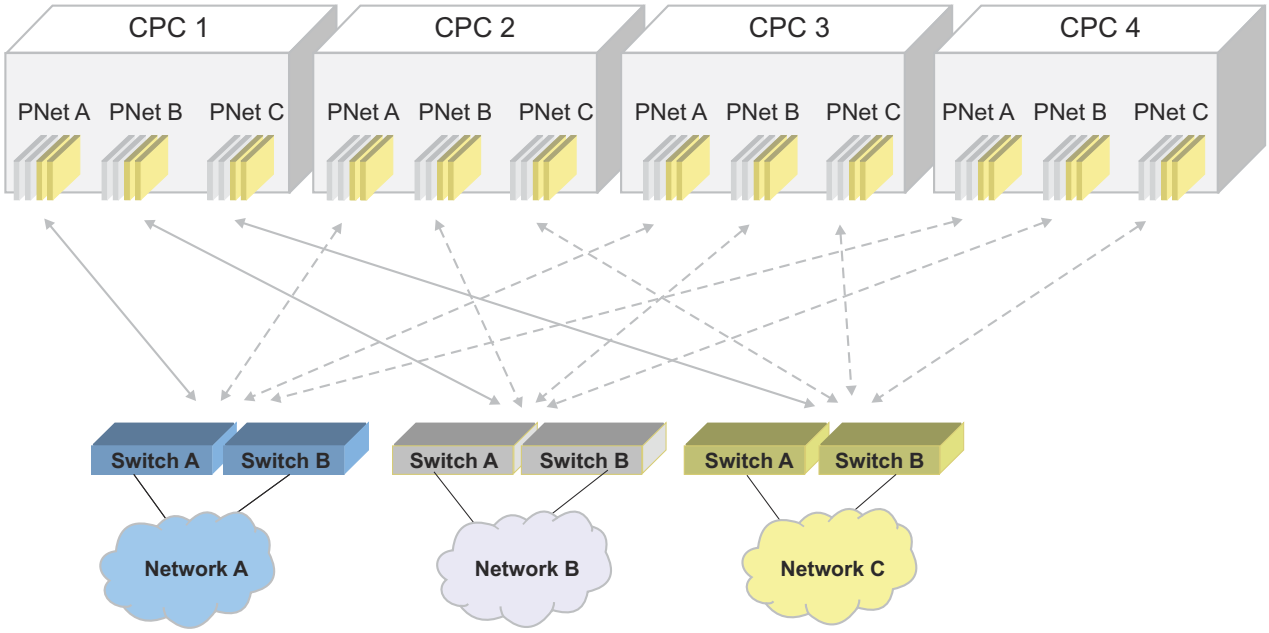


Figure 70. Physical networks

The concept of a physical network identifier (PNetID) was created to simplify this physical network configuration task. With the convention of a PNetID, you can define a value to represent the ID or name of your physical layer 2 LAN fabric or physical broadcast domain. The System z physical ports that are to be connected to the associated physical networks are then logically associated with their respective PNetIDs. The PNetID then becomes an attribute of the physical port of the feature or adapter, describing how this feature or adapter is physically connected to your data center network. You can specify the PNetID in a single step within the hardware configuration definition (HCD), enabling all operating systems of all associated CPCs to dynamically learn and use this definition.

You can use virtual LANs (VLANs) to logically separate a physical network. If you configure multiple PNetIDs for SMC-R or SMC-D, then you must ensure that each VLAN or subnet in your configuration does not span more than one PNetID.

Rules:

- The physical network that a PNetID represents can include multiple subnets, but each subnet must be completely contained within a specific PNetID.

- The physical network that a PNetID represents can include multiple VLANs, but VLANs do not span PNetIDs even if they have the same VLAN number. For example VLAN 400 on PNet A is not the same VLAN as VLAN 400 on PNet B.

SMC processing requires the use of subnet masks. For more information, see “Configuring Shared Memory Communications over RDMA” on page 530 and “Configuring Shared Memory Communications - Direct Memory Access” on page 532.

For more information about the HCD, see z/OS HCD Planning and z/OS HCD User's Guide.

SMC-R physical network considerations

A physically separate "RoCE Express" feature is provided to use RDMA over Converged Ethernet (RoCE) on Z® . This feature is used with the existing Ethernet connectivity that OSA provides. The "RoCE Express" feature provides access to the same physical Ethernet fabric that is used for traditional IP connectivity. For more information about the Ethernet switch requirements for RoCE, see “Setting up the environment for Shared Memory Communications over RDMA” on page 529 and *IBM Z® Planning for Fiber Optic Links*.

The operating systems must logically group the associated physical ports of both the "RoCE Express" and OSA adapters based on their required physical connectivity. Each central processor complex (CPC) connects to a physical network by using both OSA and "RoCE Express" ports. You can use two "RoCE Express" ports at most to connect to a physical network at a given time, but you can use as many OSA adapters as necessary for your network bandwidth or usage requirements. An example of this logical grouping, using two OSA adapters and two"RoCE Express" features, is shown in Figure 70 on page 520.

One TCP/IP stack can define up to 16 Peripheral Component Interconnect Express (PCIe) function ID (PFID) values. Each PFID value must match a FID value configured in the hardware configuration definition (HCD).

- In a dedicated RoCE environment, each PFID represents a unique PCHID definition of a 10 GbE RoCE Express feature, and only one of the two 10 GbE RoCE Express ports for the feature can be used at a time.
- In a shared RoCE environment, each PFID represents a virtual function (VF) usage of a 10 GbE RoCE Express feature or a RoCE Express2 feature, and multiple PFID values can be associated with the same physical feature and port.

To match the "RoCE Express" features with the correct OSA SMC-R capable adapters, you must define a PNetID value for both the "RoCE Express" interface (physical port) and the corresponding OSA adapters (physical port) within the HCD. The OSA ports correspond to the stack IPAQENET and IPAQENET6 interfaces. VTAM and the TCP/IP stack then dynamically learn the PNetIDs for the "RoCE Express" interface and the OSA interfaces when the "RoCE Express" interface or the OSD interface is started. The "RoCE Express" interface is associated with only SMC-R capable OSA interfaces that have the same PNetID value defined.

Guideline: The TCP/IP stack does not validate the layer 2 physical network topology or broadcast domain. PNetIDs are values that you assign, and the operating systems learn and use these assigned values but cannot validate them within the Ethernet switched fabric. Therefore, the operating system does not assure or enforce any physical network separation or isolation across different physical networks. To physically isolate unique physical networks, you must ensure that traffic on PNet A cannot reach hosts on PNet B.

SMC-D Physical Network considerations

The ISM device does not connect to the same Ethernet fabric that the SMC-D capable OSA interfaces use. However, the operating systems must still logically group the ISM device with the OSA or HiperSockets interfaces based on their required physical connectivity. When z/OS Communications Server activates the first SMC-D capable interface for a given physical network, it also attempts to activate an ISM interface for that same physical network. The Peripheral Component Interconnect Express (PCIe) function ID (PFID) value that represents the ISM device is obtained during activation processing. The possible PFID values to use for ISM devices are configured in the hardware configuration definition (HCD).

To match the ISM device with the correct SMC-D capable OSA or HiperSockets adapters, you must define a PNetID value for both the ISM device and the corresponding OSA or HiperSockets adapter within the HCD. The OSA ports correspond to the stack IPAQENET and IPAQENET6 interfaces, and the HiperSockets adapters correspond to the stack IPAQIDIO and IPAQIDIO6 interfaces. VTAM and the TCP/IP stack dynamically obtain the PNetIDs for the ISM, OSA, and HiperSockets interfaces when the interface is started. The ISM interface is associated with only SMC-D capable OSA or HiperSockets interfaces that have the same PNetID value defined.

Restriction: The same physical network cannot be used for OSA and HiperSockets interfaces. If you code the same PNetID for OSA and HiperSockets interfaces for SMC-D communications, the first interface to become fully active will be enabled for SMC-D communications and the second interface will not be able to participate in SMC-D communications.

The requirements for PNetIDs for ISM have changed for SMC Version 2. For details of the SMC-Dv2 PNetID requirements for ISM, see "SMC-Dv2 and ISM VCHID Association using PNetIDs" on page 541.

SMC-R high availability considerations

Shared Memory Communications over RDMA (SMC-R) enables high-speed peer-to-peer connections over the RDMA over Converged Ethernet (RoCE) fabric between reliable connected queue pairs (RC QPs). SMC-R defines the RC QPs as an SMC-R link, and SMC-R links are logically grouped into SMC-R link groups. For more information, see "SMC-R links" on page 510 and "SMC-R link groups" on page 512.

"RoCE Express" features at each host are required for SMC-R communications. After a TCP connection dynamically and successfully switches to SMC-R, it cannot revert to standard TCP/IP communications. Therefore, to achieve network high availability for SMC-R, it is critical to provide redundant physical network connectivity.

If the underlying "RoCE Express" interface or the associated network hardware fails, the z/OS host provides dynamic failover processing that transparently moves the TCP connections from the SMC-R links that are using the failed "RoCE Express" interface to another SMC-R link in the link group. If no other SMC-R link in the link group is available at the time of failure, the TCP connections are lost. To have a second redundant SMC-R link within a link group, two "RoCE Express" interfaces must be defined and active.

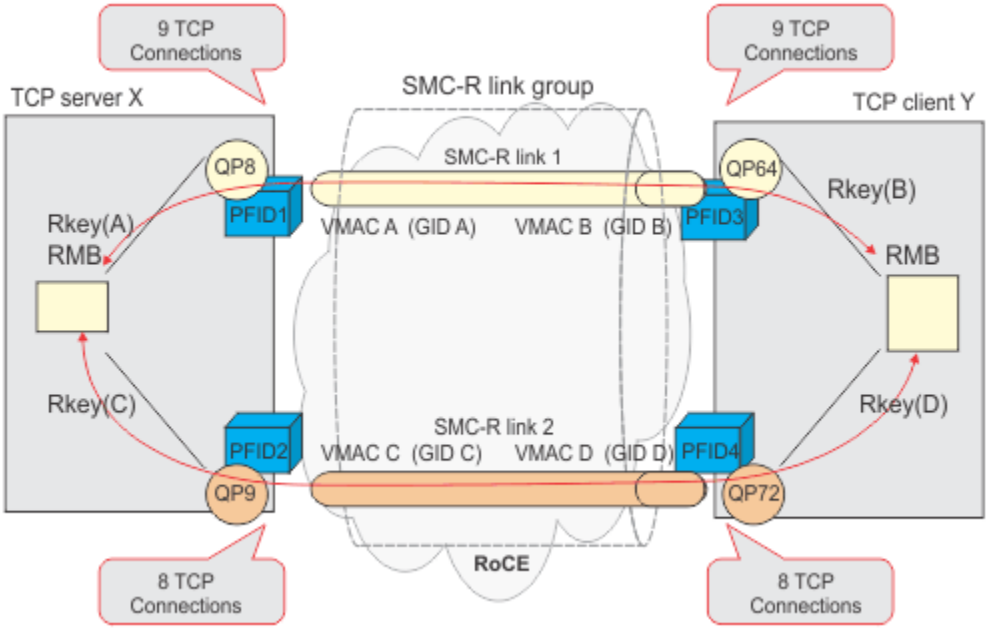


Figure 71. Redundant SMC-R links in an SMC-R link group

If the "RoCE Express" interfaces operate in a shared RoCE environment , an SMC-R link group might be considered redundant, even though the "RoCE Express" interfaces associated with SMC-R links use the same physical "RoCE Express" feature.

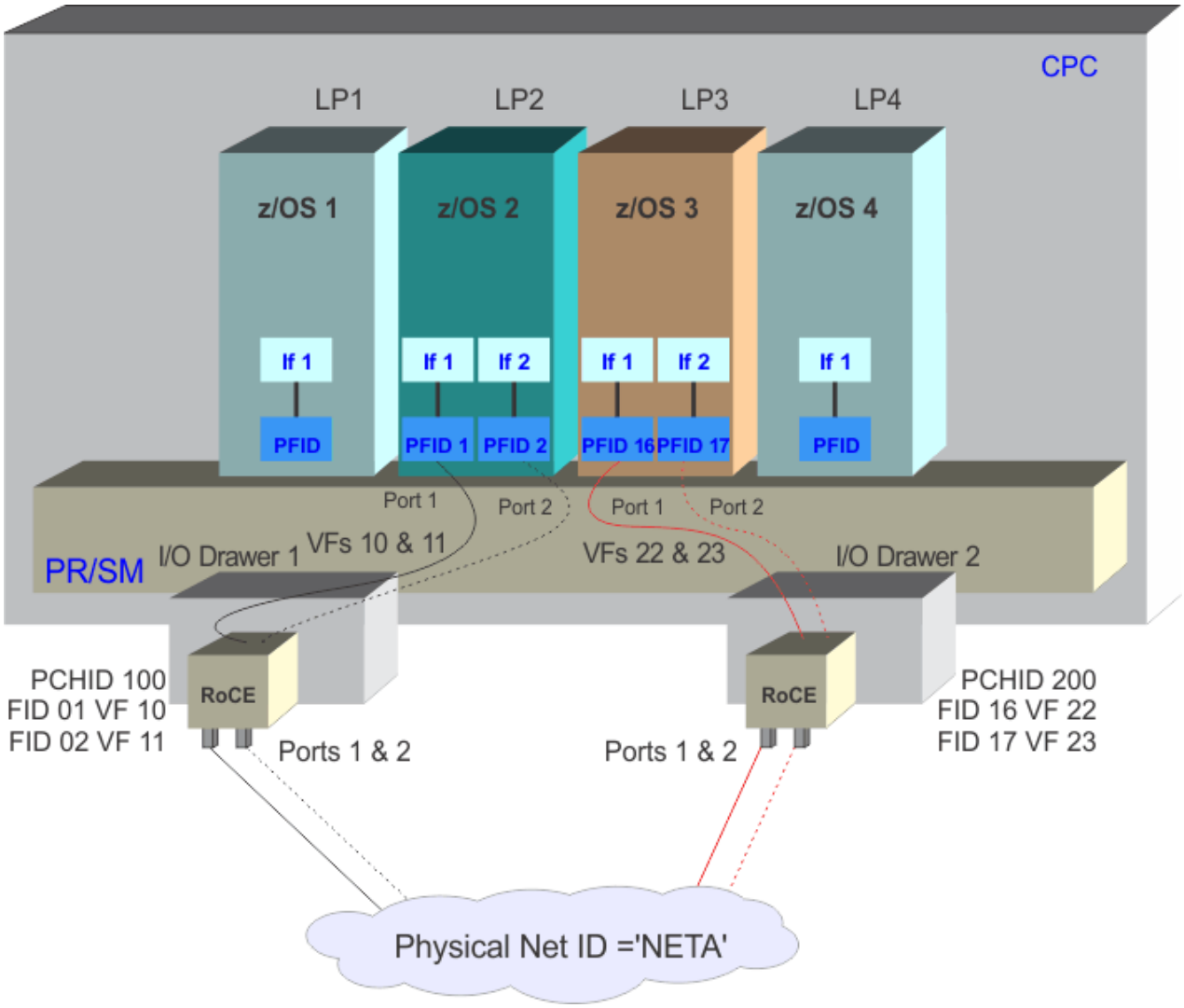


Figure 72. Misleading full redundancy configuration in a shared RoCE environment

For instance, in Figure 72 on page 523, z/OS 2 has multiple PFID values defined, but the PFID values represent different ports on the same "RoCE Express" feature. When TCP connections that use SMC-R are established in this configuration, an SMC-R link group, with two SMC-R links, is created. The two SMC-R links make this SMC-R link group appear to have full redundancy, but an failure involving the "RoCE Express" feature will result in failures of both PFIDs and all the associated interfaces. This in turn will cause failures for both SMC-R links within the SMC-R link group. As a result, dynamic failover processing will not occur, and TCP connections that use those SMC-R links will fail. A configuration of this type is identified by a value of "Partial (single local PCHID, unique ports)" in Netstat Devlinks/-d reports involving the SMC-R link group. For more information, see Redundancy levels.

To ensure that a redundant path exists in a shared RoCE environment, you must design your connectivity to ensure that the PFID values used by a given TCP/IP stack represent physically different "RoCE Express" features. Two "RoCE Express" features are physically different if they are configured with different PCHID values. See Figure 62 on page 506 for an example of using physically different "RoCE Express" features in a shared RoCE environment.

As shown in Figure 71 on page 522, when both SMC-R peers have two active "RoCE Express" interfaces, TCP connections are distributed across the links. TCP connection data can use either SMC-R link, even if the TCP connection is considered to be assigned to a specific SMC-R link.

If a failure is experienced involving one SMC-R link, all the TCP connections are moved automatically to the other SMC-R link. For example, as shown in Figure 73 on page 524, when SMC-R link 2 fails, all connections are moved to SMC-R link 1. After recovery, when a new SMC-R link is established, new TCP/IP connections are moved to the new link to balance utilization of the RoCE physical resources. Existing connection might also be moved to the new link.

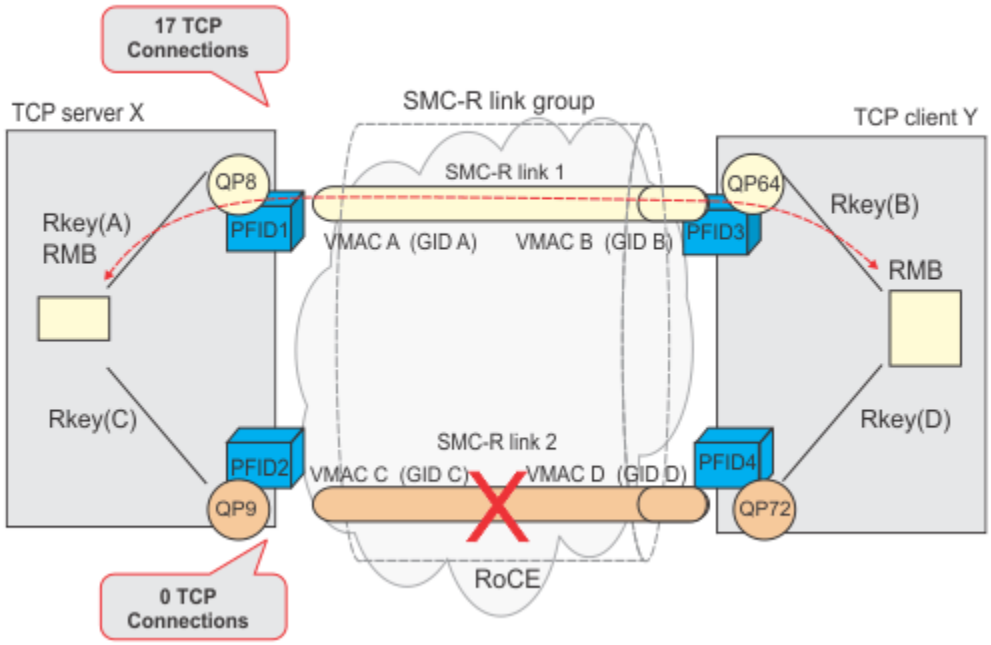


Figure 73. Failover processing within an SMC-R link group

Figure 71 on page 522 and Figure 73 on page 524 do not show the RoCE switches, but ideally, redundant physical switches are also present.

If both SMC-R peers do not have multiple active "RoCE Express" interfaces, then the SMC-R link group does not provide an ideal level of TCP connection resiliency. Figure 74 on page 525 is an example of a configuration where one peer (the server host) has two active "RoCE Express" interfaces, but the other peer (the client host) has just one. In this situation, the server still creates two SMC-R links, one per active interface, so the server can still move the TCP connections between SMC-R links if a "RoCE Express" interface fails. The client, however, cannot move the TCP connections if its "RoCE Express" interface fails because no alternative path exists. Because only one peer can provide recovery capability, this configuration has partial redundancy.

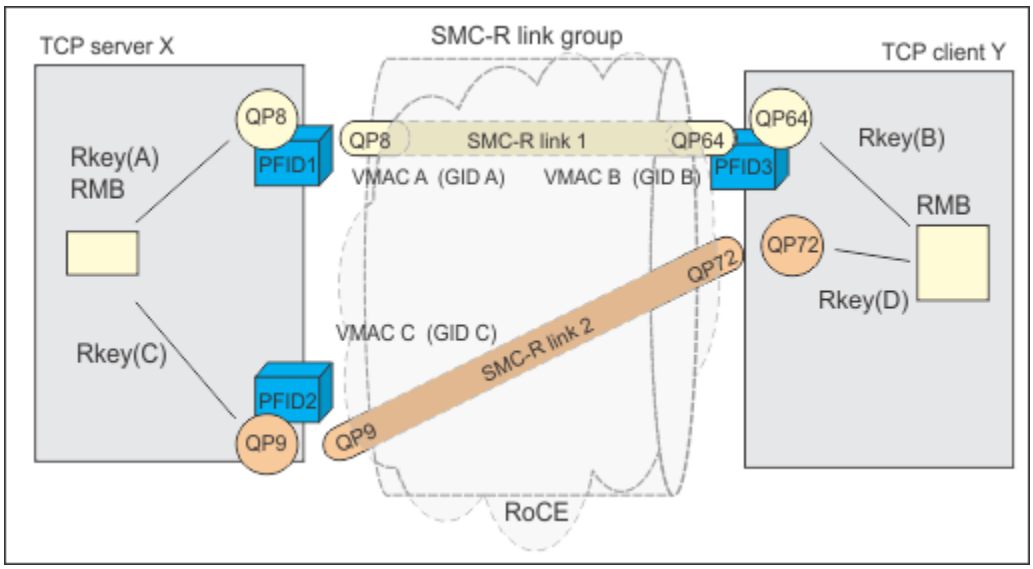


Figure 74. Partially redundant SMC-R links

If neither the server or the client has multiple active "RoCE Express" interfaces, as shown in Figure 75 on page 525, then the SMC-R link group is composed of a single SMC-R link. If a "RoCE Express" interface fails in this configuration, the TCP connections cannot be recovered or moved, so they are all lost. This type of SMC-R link is called a single link, and the configuration has no redundancy capability.

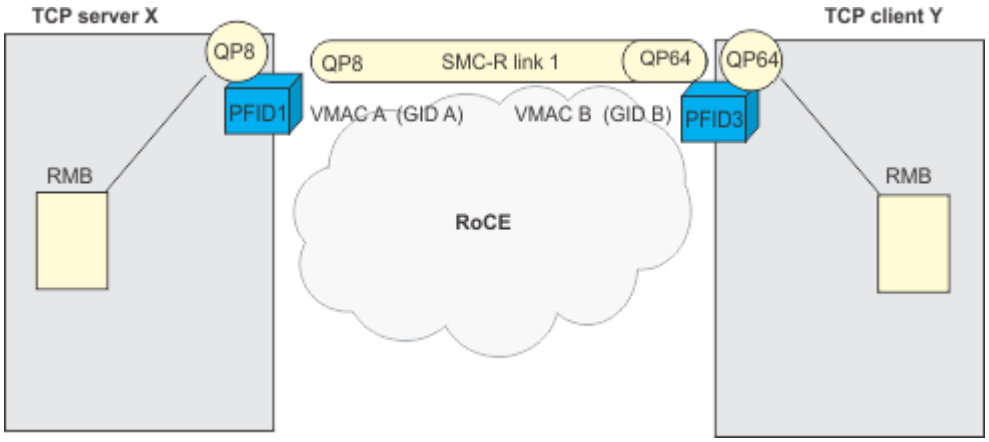


Figure 75. SMC-R link group with no redundant link

Redundancy levels

System z also provides redundant internal Peripheral Component Interconnect Express (PCIe) hardware support infrastructures for the PCIe-based "RoCE Express" features. For simplicity, the System z internal PCIe infrastructure is referred to as the *internal path*. The internal path of the "RoCE Express" feature is determined based on how the feature is plugged into the System z I/O drawers. To have full "RoCE Express" hardware redundancy on System z, each feature must have unique internal paths. For more information about the System z I/O drawer configurations, see your IBM Service representative.

A complete high availability solution, therefore, requires the following setup between two SMC-R peers:

- Two unique physical "RoCE Express" features that use unique PCHIDs (see “SMC-R high availability considerations” on page 522)
- Unique system PCIe support infrastructures, or internal paths, for the two features
- Unique physical RoCE switches

From the perspective of the local stack, the physical network topology and the internal path configuration at the remote system to the remote adapters are not visible. z/OS Communications Server can evaluate

and report a redundancy level that is based only on the known local factors. If the local stack has two unique "RoCE Express" features that have unique internal paths, then an SMC-R link group with two redundant SMC-R links is considered to have full redundancy.

Table 27 on page 526 shows the reported redundancy levels with a description of each level. The values that are listed here represent the values that are displayed for an SMC-R link group in a Netstat DEVlinks/-d report. For an example of the Netstat DEvlinks/-d report, see z/OS Communications Server: IP System Administrator's Commands.

Table 27. Redundancy levels

Redundancy level	SMC-R link group with redundant links	Unique "RoCE Express" features have unique physical internal paths	Description
Full	Yes	Yes	Full local hardware redundancy **Rule:** Hardware redundancy must be verified at each host. The internal path at the remote host is not visible to the local host and therefore is not considered.
Partial (single local internal path)	Yes	No	The local "RoCE Express" features share an internal System z PCIe adapter support infrastructure (hardware internal path). This hardware configuration provides a single point of failure, so full redundancy cannot be guaranteed.
Partial (single local PCHID, unique ports)	Yes	No	The local "RoCE Express" features use the same PCHID but unique ports. Using the same PCHID creates a single point of failure, so full redundancy cannot be guaranteed.
Partial (single local PCHID and port)	Yes	No	The local "RoCE Express" features use the same PCHID and port. Using the same PCHID and port creates a single point of failure, so full redundancy cannot be guaranteed.
Partial (single local RNIC)	No	N/A	The link group has only a single active feature on the local host, but multiple active features are available to the remote host.
Partial (single remote RNIC)	No	N/A	The link group has only a single active feature on the remote host, but multiple active features on the local host.
None (single local and remote RNIC)	No	N/A	The link group has only a single active feature on both the local and the remote host.

A "RoCE Express" interface that is associated with an SMC-R capable interface because it has the same physical network ID is referred to as an *associated RNIC interface*. More than two "RoCE Express" interfaces can be defined with the same physical network ID, but the TCP/IP stack creates SMC-R link groups that use no more than two associated RNIC interfaces at any particular time. The "RoCE Express" interfaces are considered to be associated RNIC interfaces for IPAQENET and IPAQENET6 interfaces that match all of the following characteristics:

- The interfaces are active.
- The interfaces are defined by the INTERFACE statement with the OSD channel path ID type (CHPIDTYPE OSD).
- The interfaces are enabled for SMC-R communications.
- The interfaces have matching PNetID values.

Associated RNIC interfaces are displayed in the Netstat DEvlinks/-d OSD report. For an example of the Netstat DEvlinks/-d report, see z/OS Communications Server: IP System Administrator's Commands.

Any additional "RoCE Express" interfaces that have the matching PNetID are started, but they are not used to provide for added link level load-balancing purposes. Instead, the extra "RoCE Express" interfaces are held in reserve for use if one of the associated RNIC interfaces fails.

For instance, in Figure 73 on page 524, if "RoCE Express" interface 2 (shown as PFID2) on the server host fails, the TCP connections that were using SMC-R link 2 across interface 2 are switched to SMC-R link 1. The SMC-R link group loses its level of full link redundancy because only SMC-R link 1 is active. However, if another "RoCE Express" interface, call it PFID 5, were active on the server host, and PFID 5 had the same PNetID value as PFID 1 and PFID 2, the server can immediately activate new SMC-R links across PFID 5 to the client host to reestablish full link redundancy. If PFID 5 and PFID 1 have unique physical paths, then full redundancy is also restored. This new SMC-R link is used for TCP connections within the link group. If PFID 2 recovers, it now serves as a standby PFID and can be used if either PFID 1 or PFID 5 fails.

You can also use extra PFIDs for planned outages, such as to schedule an upgrade to the "RoCE Express" features.

System and network requirements for Shared Memory Communications

Depending on what type of Shared Memory Communications (SMC) you will use, the system requirements are different. In addition, SMC-R imposes additional network requirements.

If you will use both SMC-R and SMC-D, you must ensure that the system and network requirements for each type of processing are met.

System requirements for SMC-R in a dedicated RoCE environment

You need to ensure that your system meets the requirements to use SMC-R with 10 GbE RoCE Express features operating in a dedicated RoCE environment.

A z/OS image must be z/OS Version 2 Release 1 or later to use Shared Memory Communications over RDMA (SMC-R) with 10 GbE RoCE Express features operating in a dedicated RoCE environment.

SMC-R requires RDMA over Converged Ethernet (RoCE) hardware and firmware support. The following minimum hardware requirements must be met to use SMC-R:

- You must have an IBM zEnterprise EC12 (zEC12) with driver 15, or an IBM zEnterprise BC12 (zBC12).
- You must have one or more IBM 10 GbE RoCE Express features.

  10 GbE RoCE Express features are dual ports with short range (SR) optics and dedicated to a single LPAR image.

  Guideline: Provide two 10 GbE RoCE Express features per z/OS image per unique physical network. For more information, see “RoCE network high availability” on page 528.
- You must have System z OSA-Express for traditional Ethernet LAN connectivity using CHPID type OSD.

  SMC-R does not impose any specific OSA requirements.
- You must have standard 10 GbE switches.

System requirements for SMC-R in a shared RoCE environment

You need to ensure that your system meets the requirements to use Shared Memory Communications over RDMA (SMC-R) with "RoCE Express" features operating in a shared RoCE environment.

To use SMC-R with 10 GbE RoCE Express features operating in a shared RoCE environment, the minimum software requirement must be z/OS Version 2 Release 1 with APARs OA44576 and PI12223 applied.

To use SMC-R with RoCE Express2 features, the minimum software requirements are:

- z/OS Version 2 Release 1 with APARs OA51949 and PI75199 applied
- z/OS Version 2 Release 2 with APARs OA51950 and PI75200 applied

SMC-R requires RDMA over Converged Ethernet (RoCE) hardware and firmware support. The following minimum hardware requirements must be met to use SMC-R:

- If you use 10 GbE RoCE Express features, you must have IBM z13 (z13) or later systems.
- If you use RoCE Express2 features, you must have IBM z14 or later systems
- You must have one or more IBM 10 GbE RoCE Express or RoCE Express2 features.

  "RoCE Express" features are dual ports with short range (SR) optics and can be shared across multiple operating systems images or TCP/IP stacks in a central processor complex (CPC).

  Guideline: Provide two "RoCE Express" features per unique physical network. For more information, see “RoCE network high availability” on page 528.
- You must have System z OSA-Express for traditional Ethernet LAN connectivity.

  SMC-R does not impose any specific OSA requirements.
- You must have standard 10 GbE or 25 GbE switches.
- If you configure more than 24 Peripheral Component Interconnect Express (PCIe) devices, you must configure the IEASYSxx LFAREA parameter. The 24 PCIe devices include all z/OS Communications Server PCIe devices and other z/OS PCIe devices. In z/OS Communications Server, you can configure the following PCIe devices:
  - IBM 10 GbE RoCE Express features
  - RoCE Express2 features
  - Internal shared memory (ISM) devices

  For more information about specifying the IEASYSxx LFAREA parameter, see z/OS MVS Initialization and Tuning Guide.

Network requirements for SMC-R

You need to ensure that your system meets the network requirements to use SMC-R with "RoCE Express" features.

z/OS Communications Server supports connectivity to multiple, distinct layer 2 networks through unique physical LANs. Each unique physical network is identified by existing Ethernet standards that are based on the physical layer 2 broadcast domain. You can define a physical network ID (PNetID) for each physical network. For more information, see “SMC-R physical network considerations” on page 521.

For hosts to communicate by using SMC-R, they must connect directly to the same Ethernet layer 2 LAN network. If VLANs are in use, each host must also have access to the same VLAN. For more information, see “SMC-R VLANID usage for the 10 GbE RoCE Express feature” on page 519.

There are restrictions on the physical distances that can be used to route RDMA frames. To understand these distance specifications and limitations, see *IBM z Systems Planning for Fiber Optic Links*.

RoCE network high availability

Because RoCE connections do not use IP routing and the RDMA connections to remote hosts are direct point-to-point connections that use reliable connected queue pairs (RC QPs), there is no concept of an alternative IP route to the peer. SMC-R connectivity is possible with a single "RoCE Express" feature, but a loss in that single feature means that the associated TCP connections and workloads are disrupted. Therefore, redundant "RoCE Express" features on both the local and remote hosts are required to achieve network high availability with SMC-R. If your TCP workloads require high availability, redundant "RoCE Express" features and redundant Ethernet switches are required. The SMC-R protocol actively uses both features, rather than using one feature with the other in standby mode. For more information, see “SMC-R high availability considerations” on page 522.

"RoCE Express" features also have redundant internal PCIe support structures, or PCIe internal paths, as described in “Redundancy levels” on page 525. To avoid another single point of failure, install each "RoCE Express" feature that is managed by the same operating system with a unique internal path. For

more information about how to install a "RoCE Express" feature to achieve full redundancy, see your IBM service representative.

RoCE bandwidth

The "RoCE Express" features provide 10 GbE ports. When redundant features are defined, SMC-R link groups can be formed by using both features, resulting in a 20 GbE logical pipe to each physical network. z/OS Communications Server uses only two features within a link group at any particular time.

System requirements for SMC-D

You need to ensure that your system meets the system requirements to use SMC-D.

To use Shared Memory Communications - Direct Memory Access (SMC-D), the minimum software requirement is z/OS Version 2 Release 2 with APARs OA48411 and PI45028 applied, and the minimum hardware requirement is IBM z13® GA2 or later systems.

If you configure more than 24 Peripheral Component Interconnect Express (PCIe) devices, you must configure the IEASYSxx LFAREA parameter. The 24 PCIe devices include all z/OS Communications Server PCIe devices and other z/OS PCIe devices. In z/OS Communications Server, you can configure the following PCIe devices:

- 10 GbE RoCE Express features
- RoCE Express2 features
- Internal shared memory (ISM) devices

For more information about specifying the IEASYSxx LFAREA parameter, see z/OS MVS Initialization and Tuning Guide.

Setting up the environment for Shared Memory Communications over RDMA

Before you configure Shared Memory Communications over RDMA (SMC-R), follow these steps to ensure that other components are configured.

Before you begin

Review “Configuration considerations for Shared Memory Communications” on page 518.

Procedure

Perform the following steps to prepare to use SMC-R:

1. Install and configure the "RoCE Express" features in the hardware configuration definition (HCD).

   Logical partition (LPAR) access lists must be provided for the "RoCE Express" features.
2. Assign physical network ID (PNetID) values, and configure the values in the HCD for both the "RoCE Express" ports and any OSA-Express devices that will use the "RoCE Express" ports for SMC-R communications.
3. Provide for redundant system PCIe internal paths for the defined "RoCE Express" features.

   For more information, see “Redundancy levels” on page 525.
4. Configure Ethernet switches for RDMA functionality.

   RDMA processing requires standard 10 GbE switch support, and distance limitations might exist. Enable the global pause frame (a standard Ethernet switch feature for Ethernet flow control that is described in the IEEE 802.3x standard) on the switch. You might also need to configure the switch to indicate whether you use VLANs.

   For more information, see “Network requirements for SMC-R” on page 528.

Configuring Shared Memory Communications over RDMA

Use these steps to configure and begin to use Shared Memory Communications over RDMA (SMC-R).

Before you begin

See "Setting up the environment for Shared Memory Communications over RDMA" on page 529.

Procedure

Perform the following steps to configure SMC-R:

1. If necessary, convert IPv4 IPAQENET DEVICE, LINK, and HOME definitions to INTERFACE definitions.

   SMC-R processing is provided only for OSD interfaces configured with INTERFACE definitions. For more information about converting IPv4 IPAQENET DEVICE, LINK, and HOME definitions to INTERFACE definitions, see "Steps for converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the IPv4 IPAQENET INTERFACE statement" on page 57.
2. Configure the SMCR parameter on the GLOBALCONFIG statement in the TCP/IP profile.

   The SMCR parameter includes the following subparameters:

   - PFID specifies the PCI Express (PCIe) function ID (PFID) value for a "RoCE Express" feature that this stack uses.

     You must code at least one PFID subparameter for this stack to use SMC-R, and two PFIDs per PNetID per stack for redundancy.

     – When the 10 GbE RoCE Express features operate in a dedicated RoCE environment, each 10 GbE RoCE Express feature must have a unique PFID value, but each TCP/IP stack that uses the 10 GbE RoCE Express feature specifies the same PFID value.

     – When the 10 GbE RoCE Express or RoCE Express2 features operate in a shared RoCE environment, each TCP/IP stack that uses the same "RoCE Express" feature must have a unique PFID value, even if the TCP/IP stacks are defined on different LPARs.
   - PORTNUM specifies the 10 GbE RoCE Express port number to use for each PFID.

     Guideline: You do not have to code PORTNUM for a PFID representing a RoCE Express2 feature. The port number is defined for the PFID in the HCD, and VTAM and the TCP/IP stack learns the port number during PFID activation.

     Configure each PFID to use only a single port. The port number can be 1 or 2; 1 is the default port number.

     – When the 10 GbE RoCE Express features operate in a dedicated RoCE environment, either port 1 or port 2 can be used for a particular 10 GbE RoCE Express feature, but z/OS Communications Server cannot be configured to use both ports of a feature. For example, specifying `PFID 0018 PORTNUM 1` and `PFID 0018 PORTNUM 2`, even if specified on different TCP/IP stacks in the same LPAR, results in an error during 10 GbE RoCE Express activation processing for the second port that is activated.

     – When the 10 GbE RoCE Express features operate in a shared RoCE environment, both port 1 and port 2 can be used simultaneously if the ports are associated with different PFID values. For example, assuming that PFID 0018 and PFID 0019 represent the same physical 10 GbE RoCE Express feature, you can specify `PFID 0019 PORTNUM 1` and `PFID 0018 PORTNUM 2` to use both ports.
   - MTU specifies the maximum transmission unit (MTU) to be used for this PFID. The default value is 1024. For more information, see "SMC-R RoCE maximum transmission unit" on page 558.
   - FIXEDMEMORY specifies the total amount of memory, in megabytes, that can be used for the staging and remote memory buffers.

     The default value is 256 MB.
   - TCPKEEPMININTERVAL specifies the minimum time interval, in seconds, for sending keepalive probes for TCP connections that use SMC-R protocols to exchange data.

The default value is 300 seconds. For more information, see "TCP keepalive" on page 557.

The following GLOBALCONFIG statement defines two 10 GbE RoCE Express features, PFID 0018 and PFID 0019. Port 2 is used on each feature, and the maximum amount of 64-bit private storage that can be used for SMC-R communications is 200 megabytes. The default values for both TCPKEEPMININTERVAL and MTU are used.

```
GLOBALCONFIG SMCR
    PFID 0018 PORTNUM 2
    PFID 0019 PORTNUM 2
    FIXEDMEMORY 200
```

If PFID 0018 and PFID 0019 represent RoCE Express2 features, you do not need to specify PORTNUM on the GLOBALCONFIG definition.

```
GLOBALCONFIG SMCR
PFID 0018
PFID 0019
FIXEDMEMORY 200
```

For more information about these and other SMCR subparameters on the GLOBALCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

3. (Optional) Configure the SMCR parameter on the IPAQENET and IPAQENET6 INTERFACE statements with the OSD channel path ID type (CHPIDTYPE OSD).

   Tip: SMCR is the default setting on the IPAQENET and IPAQENET6 INTERFACE statements for the OSD CHPID type.

   Guideline: If you enable Multipath and the equal-cost interfaces are associated with different IP subnets, enabling SMC for some of, but not all, the interfaces can cause unpredictable SMC usage. You must specify either SMCR or NOSMCR on all equal-cost interfaces.

4. Associate the interfaces with the appropriate subnet or prefix.

   - For an IPv4 interface to be eligible for SMC-R, you must configure a nonzero subnet mask on the INTERFACE statement in the TCP/IP profile.

     Result: SMC-R is used only between peers whose interfaces have the same subnet value.

   - For an IPv6 interface to be eligible for SMC-R, the interface must have at least one prefix that is associated with it.

     Rule: A prefix can be associated to an IPv6 interface in any of these ways:

     – A prefix received on a router advertisement message from an attached router
     – A prefix that is configured in OMPROUTE by using the PREFIX parameter on the IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, or IPV6_INTERFACE statement
     – A direct static prefix route that is configured over the interface on a ROUTE statement in a BEGINROUTES block in the TCP/IP profile

     Result: SMC-R is used only between peers whose IPv6 interfaces have at least one prefix in common.

5. (Optional) If you are using VLANs for your SMC-R communications, configure the VLANID parameter on the IPAQENET and IPAQENET6 INTERFACE statements for the OSD CHPID type.

   For more information, see "VLANID considerations" on page 518.

6. (Optional) If you have a server application that primarily uses short-lived TCP connections, you might want to avoid SMC-R rendezvous processing for TCP connections that are using that server port.

   Configure NOSMC on the PORT or PORTRANGE statement for the server port or ports that this server application uses. For more information, see z/OS Communications Server: IP Configuration Reference.

7. Start the IPAQENET and IPAQENET6 interfaces.

   When the first SMC-R capable OSD interface becomes active, z/OS Communications Server automatically starts all PFIDs that are defined in the GLOBALCONFIG statement, and associates the 10 GbE RoCE Express interfaces with the OSD interfaces that have matching physical network IDs (PNet

IDs). For more information about PNet IDs, see “SMC-R physical network considerations” on page 521.

What to do next

For information about how SMC-R interacts with other functions, see “SMC interactions with other z/OS Communications Server functions” on page 555.

For information about managing SMC-R communications, see “Managing SMC communications” on page 558.

Setting up the environment for Shared Memory Communications - Direct Memory Access

Before you configure Shared Memory Communications - Direct Memory Access (SMC-D), follow these steps to ensure that other components are configured.

Before you begin

Review “Configuration considerations for Shared Memory Communications” on page 518.

Procedure

Assign physical network ID (PNetID) values, and configure the values in the HCD for both the ISM devices and any OSA or HiperSockets devices that will use the ISM device for SMC-D communications.

Configuring Shared Memory Communications - Direct Memory Access

Use these steps to configure and begin to use Shared Memory Communications - Direct Memory Access (SMC-D).

Before you begin

See “Setting up the environment for Shared Memory Communications - Direct Memory Access” on page 532.

Procedure

Perform the following steps to configure SMC-D:

1. If you are using IPv4 IPAQENET DEVICE, LINK, and HOME definitions, convert them to INTERFACE definitions.

   SMC-D processing is provided for only OSD interfaces that are configured with INTERFACE definitions. For more information about converting IPv4 IPAQENET DEVICE, LINK, and HOME definitions to INTERFACE definitions, see “Steps for converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the IPv4 IPAQENET INTERFACE statement” on page 57.
2. If you are using IPv4 IPAQIDIO DEVICE, LINK, and HOME definitions, convert them to INTERFACE definitions.

   SMC-D processing is provided for only HiperSockets interfaces that are configured with INTERFACE definitions. For more information about converting IPv4 IPAQIDIO DEVICE, LINK, and HOME definitions to INTERFACE definitions, see “Steps for converting from IPv4 IPAQIDIO DEVICE, LINK, and HOME definitions to the IPv4 IPAQIDIO INTERFACE statement” on page 75.
3. Configure the SMCD parameter on the GLOBALCONFIG statement in the TCP/IP profile.

   The SMCD parameter includes the following subparameters:

   - FIXEDMEMORY specifies the total amount of memory, in megabytes, that can be used for the direct memory buffers. The default value is 256 MB.

- TCPKEEPMININTERVAL specifies the minimum time interval, in seconds, for sending keepalive probes for TCP connections that use SMC-D protocols to exchange data. The default value is 300 seconds. For more information, see "TCP keepalive" on page 557.

For more information about SMCD subparameters on the GLOBALCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

4. (Optional) Configure the SMCD parameter on the IPAQENET and IPAQENET6 INTERFACE statements with the OSD channel path ID type (CHPIDTYPE OSD).

   Tip: SMCD is the default setting on the IPAQENET and IPAQENET6 INTERFACE statements for the OSD CHPID type.

   Guideline: If you enable Multipath and the equal-cost interfaces are associated with different IP subnets, enabling SMC for some of, but not all, the interfaces can cause unpredictable SMC usage. You must specify either SMCD or NOSMCD on all equal-cost interfaces.

5. (Optional) Configure the SMCD parameter on the IPAQIDIO and IPAQIDIO6 INTERFACE statements.

   Tip: SMCD is the default setting on the IPAQENET and IPAQENET6 INTERFACE statements.

   Guideline: If you enable Multipath and the equal-cost interfaces are associated with different IP subnets, enabling SMC for some of, but not all, the interfaces can cause unpredictable SMC usage. You must specify either SMCD or NOSMCD on all equal-cost interfaces.

6. Associate the interfaces with the appropriate subnet or prefix.

   - For an IPv4 interface to be eligible for SMC-D, you must configure a nonzero subnet mask on the INTERFACE statement in the TCP/IP profile.

     Result: SMC-D is used only between peers whose interfaces have the same subnet value. SMC-Dv2 supports SMC-D across multiple IP subnets. For more information, see "Shared Memory Communications multiple IP subnet support (SMCv2)" on page 534.

   - For an IPv6 interface to be eligible for SMC-D, the interface must have at least one prefix that is associated with it.

     Rule: A prefix can be associated to an IPv6 interface in any of these ways:

     – A prefix that is received on a router advertisement message from an attached router
     – A prefix that is configured in OMPROUTE by using the PREFIX parameter on the IPV6_OSPF_INTERFACE, IPV6_RIP_INTERFACE, or IPV6_INTERFACE statement
     – A direct static prefix route that is configured over the interface on a ROUTE statement in a BEGINROUTES block in the TCP/IP profile

     Result: SMC-D is used only between peers whose IPv6 interfaces have at least one prefix in common. SMC-Dv2 supports SMC-D across multiple IPv6 prefix values. For more information, see "Shared Memory Communications multiple IP subnet support (SMCv2)" on page 534.

7. If you use VLANs for your SMC-D communications, configure the VLANID parameter on the IPAQENET and IPAQENET6 INTERFACE statements for the OSD CHPID type and on the IPAQIDIO and IPAQIDIO6 INTERFACE statements.

   For more information, see "VLANID considerations" on page 518.

8. If you have a server application that primarily uses short-lived TCP connections, you might avoid SMC-D rendezvous processing for TCP connections that are using that server port.

   Configure the NOSMC parameter on the PORT or PORTRANGE statement for the server port or ports that this server application uses. For more information, see z/OS Communications Server: IP Configuration Reference.

9. Start the IPAQENET and IPAQENET6 interfaces, or the IPAQIDIO and IPAQIDIO6 interfaces.

   When the first SMC-D capable OSD or HiperSockets interface becomes active for a given physical network ID (PNetID), z/OS Communications Server automatically looks for an available ISM device with a PNetID value that matches the PNetID value of the OSD or HiperSockets interface. If a matching ISM device is available, VTAM activates the ISM device and provides the associated PFID to the TCP/IP stack. For more information about PNetIDs, see "Physical network considerations" on page 520. For

details of the SMC-Dv2 PNetID requirements for ISM, see “SMC-Dv2 and ISM VCHID Association using PNetIDs” on page 541.

What to do next

For information about how SMC interacts with other functions, see “SMC interactions with other z/OS Communications Server functions” on page 555.

For information about managing SMC communications, see “Managing SMC communications” on page 558.

Shared Memory Communications multiple IP subnet support (SMCv2)

The initial version of SMC was limited to TCP/IP connections over the same layer 2 network and therefore was not routable across multiple IP subnets. The associated TCP/IP connection was limited to hosts within a single IP subnet requiring the hosts to have direct access to the same physical layer 2 network (i.e. same Ethernet LAN over a single VLAN ID). The scope of eligible TCP/IP connections for SMC was limited to and defined by the single IP subnet.

SMC Version 2 (SMCv2) provides support for SMC over multiple IP subnets for both SMC-D and SMC-R and is referred to as SMC-Dv2 and SMC-Rv2. SMCv2 requires updates to the underlying network technology. SMC-Dv2 requires ISMv2 and SMC-Rv2 requires RoCEv2.

In some of the following topics it will be necessary to distinguish the SMC-Dv2 or SMC-Rv2 support from the initial SMC-D or SMC-R support. For this reason, the initial version of SMC support is now referred to as SMC-Dv1 or SMC-Rv1. The SMCv2 protocol is downward compatible allowing SMCv2 hosts to continue to communicate with SMCv1 down-level hosts.

Many architectural aspects of SMCv2 apply to both SMC-Dv2 and SMC-Rv2. The common SMCv2 topics are described first, followed by the topics that are unique to SMC-Dv2 and SMC-Rv2. While SMCv2 changes the SMC connection protocol enabling multiple IP subnet support, SMCv2 does not change how actual user TCP socket data is transferred, which preserves the benefits of SMC to TCP workloads.

TCP/IP connections that require IPSec are not eligible for any form of SMC.

SMC Enterprise ID

With SMCv2, the IP subnet no longer defines the scope of eligible TCP/IP connections. With the multiple IP subnet support, the SMCv2 administrator will require a new means for defining or limiting the scope of hosts that are eligible for SMC over multiple IP subnets. The SMC Enterprise ID (EID) defines a group of hosts that are allowed to communicate using SMCv2.

SMCv2 supports the concept of allowing the SMCv2 administrator to define a single unique ID that represents a group of systems. Systems means OS instances (z/OS or Linux), potentially many systems or the group could be a smaller set of systems. For SMC-D, the systems must all reside (execute) on the same CPC.

While SMCv2 extends the reach of SMC to additional workloads, the objective of SMCv2 is optimizing workloads within your enterprise data center. SMC-Dv2 connections must execute on systems within the same CPC. SMC-Rv2 is not targeting long distance connections over the WAN. For SMC-Rv2, the physical proximity of and distance between your systems are key considerations.

In addition to Enterprise IDs (EIDs), SMCv2 offers administrators a more granular form of controlling SMC connection eligibility with SMC filters. SMC filters apply to all versions and types of SMC (SMC-Dv1, SMC-Dv2, SMC-Rv1, and SMC-Rv2). See “SMC filters” on page 539 for more details.

EID Concepts

Systems that are eligible to connect using SMCv2 must define a common EID. The systems could span many IP subnets. The IP subnet of the peer host's IP interface is not evaluated when determining SMCv2 connection eligibility.

The EID is a user defined identifier or UEID (User defined EID) that is defined once and then used (configured) among all systems within the group. A System EID or SEID is also available for SMC-Dv2 only and is described below.

The administrator would first identify the systems where SMC would be deployed across multiple IP subnets within their Enterprise. This could be a single group or it could be multiple groups of systems. SMC groups (EIDs) are not directly associated with other clustering technologies such as Sysplex. Instead, the SMC group is simply a group of heterogenous systems where SMC is to be permitted across multiple IP subnets.

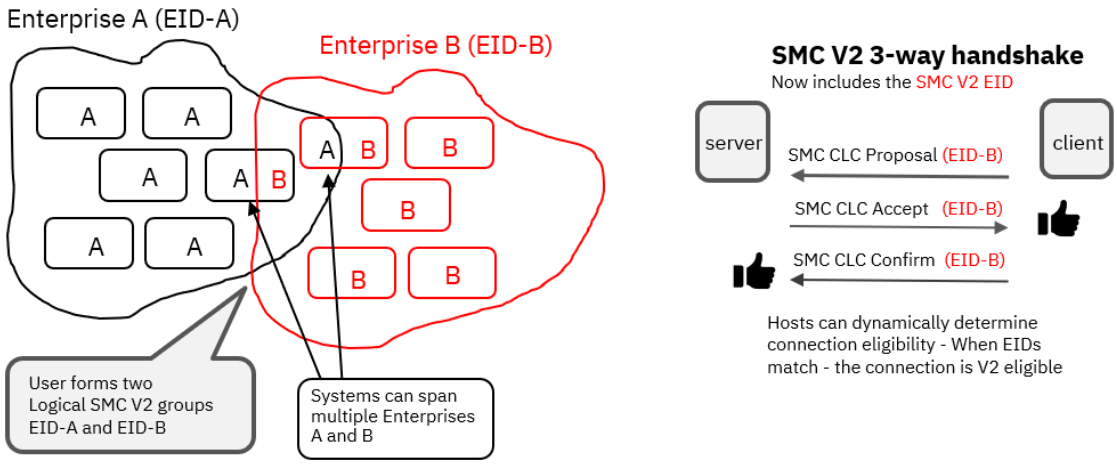


Figure 76. SMCv2 EID concepts

Each group (or groups) of systems within an Enterprise would be defined with a unique EID. The EID is an SMC configuration parameter describing a group of systems (within the Enterprise) eligible to connect over multiple IP subnets using SMCv2. The EID is not intended to represent authorization, it is not

encrypted or encoded in any way. Existing TCP connection level authentication protocols such as TLS are used to control connection level authentication.

Defining globally unique user defined EIDs is important for avoiding potential conflicts with other systems, particularly systems outside your control.

The EID attributes are summarized here:

- An EID is a single ID (per unique group of systems) representing an SMC group of systems.
- The EID applies to the entire TCP/IP stack for both IPv4 and IPv6.[4]
- The EID is not associated to specific IP interfaces, IP networks, or IP subnetworks.
- The EID is a software attribute associated with the SMCv2 protocol. The EID is not a network hardware attribute.
- The EID is exchanged within the SMCv2 connection protocol where the EIDs are dynamically evaluated during the connection setup process to determine eligibility.

EID Format

The EID is a fixed length 32-byte character data that is a user defined ID. The EID is converted to ASCII characters when used within SMC protocol messages. The UEID supports any valid combination of the valid ASCII characters (ASCII 4) described in the EID definition below with no spaces or blanks.

The valid ASCII characters allowed for the EID are upper case A-Z, 0-9, hyphen, and dot. The first character of the EID cannot be a special character and the EID cannot contain consecutive dots.

The EID format defined by the SMCv2 wire protocol is described as follows:

```
                                EID Format
EID: 32 ASCII character bytes
ASCII encoding: A-Z upper case, 0-9, and delimiters (special characters) hyphen "-" and dot ".".
```

User Defined EID (UEID)

User defined EIDs allow administrators to create meaningful user defined Enterprise IDs that are easily recognizable. This can be helpful when performing various administrative tasks when managing and monitoring data center networks. Users can establish unique enterprise wide EID naming conventions (e.g. A400-PRODCLUSTER4, CC29-DEVCLUSTER18 etc.). The hyphen and dot allows users to compose EIDs using elements of their IT infrastructure, such as Company-Loc-Group, BusinessLine-Group, Plex™-Cluster-Type, etc.

It is recommended that all 32 characters be defined by the user. When fewer than 32 characters are defined by the administrator, the EID is padded with trailing character blanks to form a full 32-byte EID. When the EID is exchanged within SMC connection setup messages, it must be fixed length 32 bytes. Once the user initially defines the EID, the same EID must be configured in each OS that will be permitted to communicate using SMCv2 over multiple IP subnets within the same SMC group. For more information about defining the UEID, see the SMCEID parameter on the TCP/IP profile GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Tip:

In use cases in which instances of z/OS are moved between different sites (e.g. DR sites), the UEID value can be sensitive to the current physical location of the site. For this case, users should consider setting UEID based on using z/OS System Symbolics.

Multiple EIDs

In some use cases, a single EID will be sufficient for your entire enterprise data center. In this use case, there is no recognized need to separate systems using SMCv2. There are also valid use cases where it would be beneficial to separate the systems by defining multiple SMC groups. For example, users might

[4] SMC-Dv2 supports both IPv4 and IPv6. SMC-Rv2 supports IPv4 only.

want to separate and control the usage based on business lines, security zones, test versus production systems or other variations. In this case, where the systems form a more homogenous grouping, a user would define each unique group across their enterprise with a unique UEID.

Typically each z/OS system would be defined with a single UEID and belong to a single SMC group. However, there are also use cases in which a z/OS system needs to reside within multiple groups supporting multiple EIDs. z/OS Communications Server allows each TCP/IP stack to define up to four user-defined EIDs. The SMCv2 protocol permits the exchange of multiple UEIDs. If a common EID is exchanged between the client and server, the TCP connection is eligible to use SMCv2.

System EID (SEID)

SMC-Dv2 supports TCP/IP connections spanning multiple IP subnets. However, SMC-Dv2 communications are restricted to within the same CPC. Hosts must also share a common ISM VCHID. For example, two z/OS systems executing on the same CPC could establish a TCP/IP connection over OSA traversing one or more IP networks or IP router hops. The TCP connection (spanning multiple IP subnets) could be eligible to communicate internally using SMC-Dv2.

Given that SMC-Dv2 communications are restricted to a CPC, SMC-Dv2 also provides support for a system generated EID called a System EID (SEID). The SEID is automatically defined representing the CPC. The format of the SEID is consistent with the user defined EIDs. The SEID value is not defined by the user. SEID is a self-defined constant value generated by the OS encoded with system serial/type. There is a single unique SEID per CPC. The format and value of the SEID is based on IBM internal definitions described by the SMC-Dv2 protocol specifications. For more information about defining the SEID, see the SYSTEMEID/NOSYSTEMEID parameter on the TCP/IP profile GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Using the SEID provides benefits in terms of ease of use by reducing the SMC-Dv2 deployment time. The SEID allows users to enable and start using SMC-Dv2 within a CPC with minimal software configuration requirements. The System EID is applicable to an entire Z CPC and it is applicable to SMC-Dv2 and ISMv2 only. In many SMC-Dv2 use cases, a single EID will be sufficient. Using the System EID enables all SMC-Dv2 capable hosts within the same CPC that also specify the SEID to communicate using SMC-Dv2.

For more information about defining SMC EIDs, see the GLOBALCONFIG statement in z/OS Communications Server: IP Configuration Reference.

Result:

Users should be aware that defining UEID or SEID enables SMCv2 for the stack. Defining UEID enables both SMC-Rv2 and SMC-Dv2. Defining SEID enables only SMC-Dv2 (if not already enabled by UEID).

Using UEID or SEID

In some cases, the SEID could be too broad or too general for the various types of systems executing within the same CPC. Users who have a need for separation or isolation by creating multiple SMC groups within a CPC or who need a greater level control for other reasons have the following options:

1. Use the SEID (single EID per CPC) and then create separation of unique groups by assigning unique ISM VCHIDs to each group.
2. Use a unique UEID for each unique group allowing UEID to create the internal separation. Users who require greater levels of control can refrain from using the System EID and elect to only use UEIDs. Each unique group can either share a single ISM VCHID or each group can also use unique VCHIDs. Unique UEIDs with unique ISM VCHIDs provide the strongest form of separation.
3. Options 1 and 2 could be combined to meet your requirements for separation. Users can elect to use both UEIDs in combination with SEID.

Tip: You can verify the EIDs defined for your TCP/IP stack using the Netstat Config command. You can also determine if your TCP connections are using SMC-Dv2 and the EIDs used for the TCP connections using the Netstat All command.

SMC filters

In addition to organizing your SMCv2 eligible systems within groups with a common UEID, users also have the ability to control SMC connectivity at a more granular level. SMC filters are provided in the form of SMC Permit and Exclude filters. SMC filters are not restricted to SMCv2 connections. SMC filters apply to all versions and types of SMC (SMC-Dv1, SMC-Dv2, SMC-Rv1, and SMC-Rv2).

When the SMC filters are defined, they are applied to the remote host's IP address or IP subnet associated with the TCP/IP connection. The filters are not applied to the remote RoCEv2 IP address. The filters control all SMC connectivity by permitting or excluding connectivity by controlling the SMC support indication in the TCP SMCR options flag exchanged during the TCP/IP connection 3-way handshake. When connections are not permitted or specifically excluded by the SMC filters, the peer hosts view the TCP/IP connection as not eligible for SMC.

Use of SMC filters is optional. For some use cases, the base SMC connection eligibility criteria (IP subnet for SMCv1 and EID for SMCv2) is too broad. Users who need more granular control over the SMC connection eligibility can choose to exploit the SMC filters.

Before defining SMC filters, it is important to understand a couple of the key implications of using SMC filters.

Results:

- When adding your first or a single SMC Permit Filter, the default Permit behavior must be understood. With no SMC Permit Filters defined, by default all peer hosts are permitted to use SMC. However, once the first Permit Filter is added, only peer hosts that match the single (current) permit filter(s) will be permitted to use SMC.
- When your SMC Permit and Exclude Filters overlap, the Exclude Filter takes precedence over the Permit Filter, regardless of which filter is more specific.

Using SMC filters

Some examples of using SMC Permit Filters are provided here.

The SMC Permit Filters allow the following options:

- New SMC users to gradually increase the scope of SMC exploitation
- Existing users to assert more granular control on their current usage
- All users to enable specific hosts by targeting specific workloads
- Users to avoid exposing the SMC option to unknown hosts
- SMCv2 users who frequently move z/OS systems to different sites can combine EIDs with z/OS symbols and SMC filters to assure connections over the WAN are not eligible for SMCv2

The SMC Exclude Filters allow the following options:

- The exclusion of a specific IP address or set of IP addresses within a large IP subnet or site
- Overriding your Permit Filters
- The exclusion of a specific IP address that exhibits undesirable behavior
- The exclusion of a specific IP address on a temporary basis

Tips:

- SMC filters must be evaluated during the TCP/IP connection processing. When defining your SMC filters administrators should give this aspect consideration. To achieve the required level of SMC connection eligibility control, users should strive to use the minimum number of SMC filters.
- Users who must create SMC filters that also have a need to share the filters among several z/OS instances, or who frequently change their filters should consider using the z/OS Network Configuration Assistant (NCA) TCP/IP Profile technology to share and manage their SMC filter definitions.

Rule:

Users can configure up to a maximum of 256 Permit and 256 Exclude filters per IP version.

Dynamic selection of SMC version and SMC type

The SMC protocol uses the TCP 3-way handshake to dynamically select the SMC version and type by the TCP server. The selection is based on the client and server SMC configuration and other criteria. The TCP client can offer a single SMC version and type or up to all four SMC versions / types. The TCP server will favor SMCv2 over SMCv1 and SMC-D over SMC-R. The SMC type is dynamically selected by the server in the following order of preference:

1. SMC-Dv2
2. SMC-Dv1
3. SMC-Rv2
4. SMC-Rv1

Result:

SMCv2 is used even when client and server hosts are attached to the same IP subnet.

SMC-Dv2

SMC-Dv2 hosts must continue to be on the same CPC and have access to the same ISM VCHID. However, the associated TCP/IP connection can now span multiple IP subnets crossing multiple "IP router hops". This also means that the IP connection through OSA or HiperSockets to a network used by each peer host is not limited to a single physical layer 2 network or VLAN.

There are additional simplification advantages in using SMC-Dv2. For this reason, when SMC-Dv2 is enabled by both peers and the TCP connection does not cross multiple IP subnets, SMC-Dv2 will be preferred or used when possible.

ISMv2

IBM System Z Internal Shared Memory (ISM) technology provides the internal communications capability required for SMC-D. The initial version of SMC-D inherited the host's IP connection's (OSA or HiperSockets) layer 2 attributes, such as PNetID (physical network) and VLAN ID to form SMC-D connectivity.

SMC-Dv2 connections are no longer bound to a single IP subnet. This allows all hosts to span multiple IP subnets connecting over multiple IP hops. SMC-Dv2 connectivity is not based on or restricted to layer 2 connectivity. The System Z Internal Shared Memory (ISM) technology is updated to support this new form of ISM Layer 3 connectivity. This update is referred to as ISMv2 and is provided with IBM z15™ at GA1.5 level.

The changes in SMC-Dv2 and ISMv2 are transparent to z/VM. z/VM 7.1 or higher is required to support all versions of SMC-D.

Like SMC-Dv1, SMC-Dv2 does not build, use, or exchange IP packets. ISMv2 does not use any form of IP routing. Instead, ISMv2 creates a unique internal ISM L3 connection that is isolated from existing L2 traffic defined by the System Z ISM architecture and the SMC-Dv2 protocol.

Enabling SMC-Dv2

When an EID is defined, with either the SEID or UEID specified in the TCP/IP profile on the Global Configuration statement, then this step also enables SMC-Dv2 for this TCP/IP stack for both IPv4 and IPv6. Once SMC-Dv2 is enabled, as new TCP/IP connections are created, they become eligible for SMC-Dv2 (enables SMC-D across multiple IP subnets).

In order for a TCP/IP connection to be eligible for SMC-Dv2, the following conditions must all be true, both the TCP client and the server must:

- Support and be enabled for SMC-Dv2 (exchanged in SMC connection setup)

- Be configured with the same EID (either SEID or UEID, exchanged in the SMC connection setup)
- Execute on the same CPC that supports ISMv2 (dynamically evaluated in the SMC connection setup processing)
- Have access to a common ISM VCHID (exchanged in the SMC connection setup)
- Not be prevented from SMC connectivity due to the presence of an SMC filter
- Not require IPSec encryption (for the associated TCP/IP connection)

When connecting with SMC-Dv2, the IP topology of the associated TCP/IP connection is not relevant and is not evaluated. This means that the physical networks, the IP subnets and the VLANs are not part of the SMC-Dv2 connection eligibility criteria and are not evaluated. The TCP/IP connection will set up across the IP network potentially over multiple IP hops and the SMC-Dv2 socket application communications occur internally over ISMv2.

Tip:

Some firewalls can be configured to drop TCP options. The SMC protocol requires TCP option 254 to flow on the TCP/IP 3-way handshake. It is recommended that you verify that your firewalls permit TCP option 254 to flow. For more information about the SMC protocol, see RFC 7609.

SMC-Dv2 and ISM VCHID Association using PNetIDs

With the initial version of SMC-D, the ISM connectivity was based on the layer 2 IP connectivity provided by OSA or HiperSockets. To allow ISM to inherit the L2 attributes, the ISM VCHID was required to be associated with an IP device such as your OSA or HiperSockets physical layer 2 network. The association was accomplished by defining the same PNetID on both ISM and OSA or HS.

With SMC-Dv2, the layer 2 network requirements are no longer part of the SMC connection criteria. This key difference means that to connect with SMC-Dv2, your ISM VCHID is not required to be associated with any IP network device, meaning ISM does not require a PNetID definition.[5]

Users who plan to exploit SMC-Dv2 have the following ISM VCHID system configuration options to consider for defining their SMC-Dv2 and ISMv2 environment:

- **ISM VCHIDs with PNetID**

  Define ISM VCHIDs with a PNetID in HCD that matches your OSA or HiperSockets PNetID definitions. This method is referred to as associated ISM devices (FIDs), meaning the ISM FIDs are directly associated with an IP device (OSA or HiperSockets) and are only used by those IP devices with the same PNetID. This method is required for SMC-Dv1. In cases in which your z/OS system supports SMC-Dv2 but it will also communicate with down-level hosts (that do not support SMC-Dv2), then you will need to define your ISM VCHID with a PNetID to meet the layer 2 connectivity requirements. You might also choose to only use associated ISM devices if you would like to continue to control the traffic flowing over the various ISM devices, as in SMC-Dv1. The interface related to an associated ISM FIDs are activated when the associated OSA or HiperSockets interface is activated.

[5] The ISM PNetID could still be required to communicate with down-level hosts.

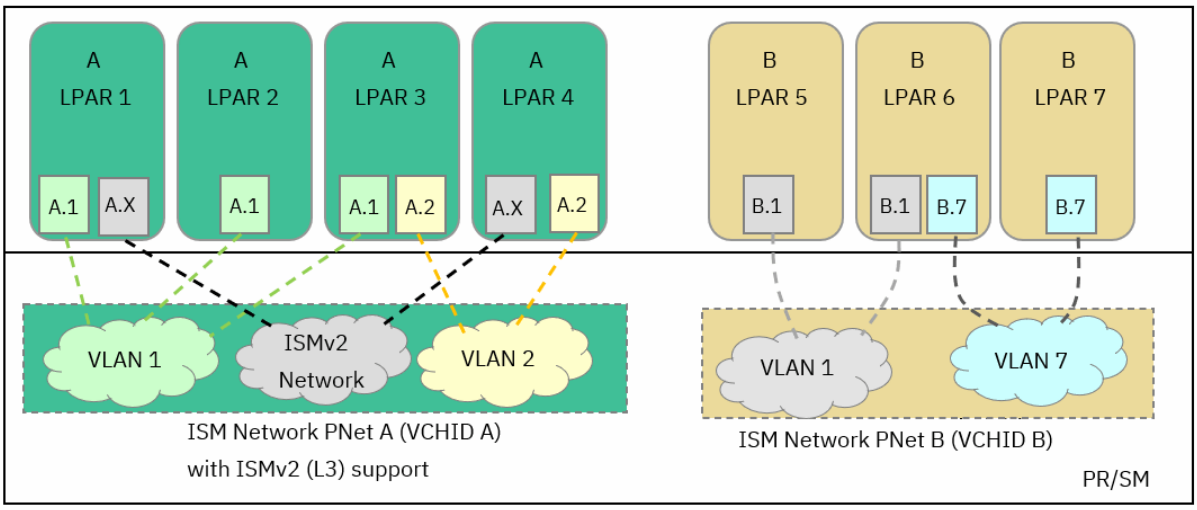


Figure 77. SMC-Dv2 and ISMv2 with PNetID

- **ISM VCHIDs without PNetID**

  Define ISM VCHIDs without a PNetID in HCD. This method is referred to as unassociated ISM devices (FIDs), meaning the ISM FIDs are not associated with any IP devices and therefore can be used by any

SMC-D eligible IP device regardless of whether a PNetID was specified for that IP device or not. Note that this also means that traffic for IP devices without a PNetID, which was not eligible for SMC-D in the past, is now eligible for SMC-Dv2. In cases in which your z/OS system supports SMC-Dv2 and will only communicate with up-level systems that also support SMC-Dv2, then you don't require an ISM PNetID. The layer 2 attributes are not associated with ISM and therefore not required to meet the SMC-Dv2 connection criteria. z/OS also supports multiple (up to four) unassociated ISM VCHIDs. Using multiple ISM VCHIDs permits the system administrator to separate and isolate unique users on unique ISM VCHIDs (e.g. unique business groups). Once the first OSA or HiperSocket interface has been started, unassociated ISM FIDs are automatically activated when the unassociated FID comes online.

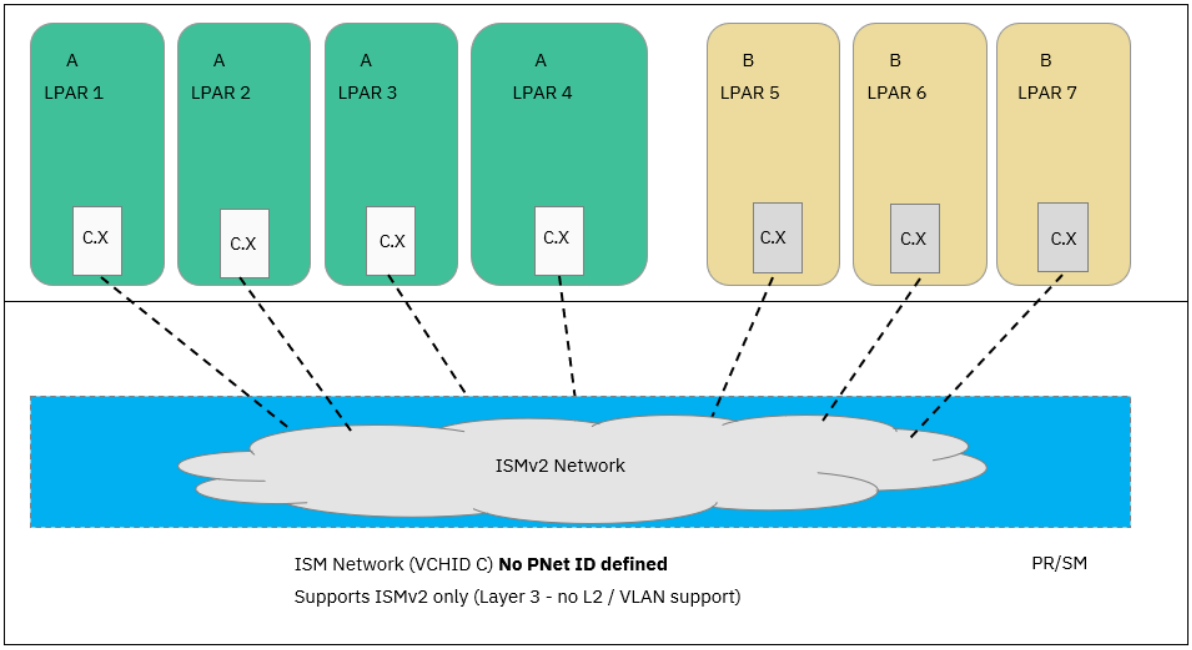


Figure 78. SMC-Dv2 and ISMv2 without PNetIDs

- **Combination of ISM VCHIDs with and without PNetID**

  Users can also combine both of the methods described above. Users can use ISM with a PNetID to communicate with down-level systems and provision a separate ISM VCHID (or multiple ISM VCHIDs)

without a PNetID used for communicating with up-level systems. Note that SMC-D eligible IP devices with a PNetID can use either ISM VCHIDs with a matching PNetID or ISM VCHIDs without a PNetID, but will prefer the ISM VCHID with the matching PNetID if the SMC partner has access to that same ISM VCHID.

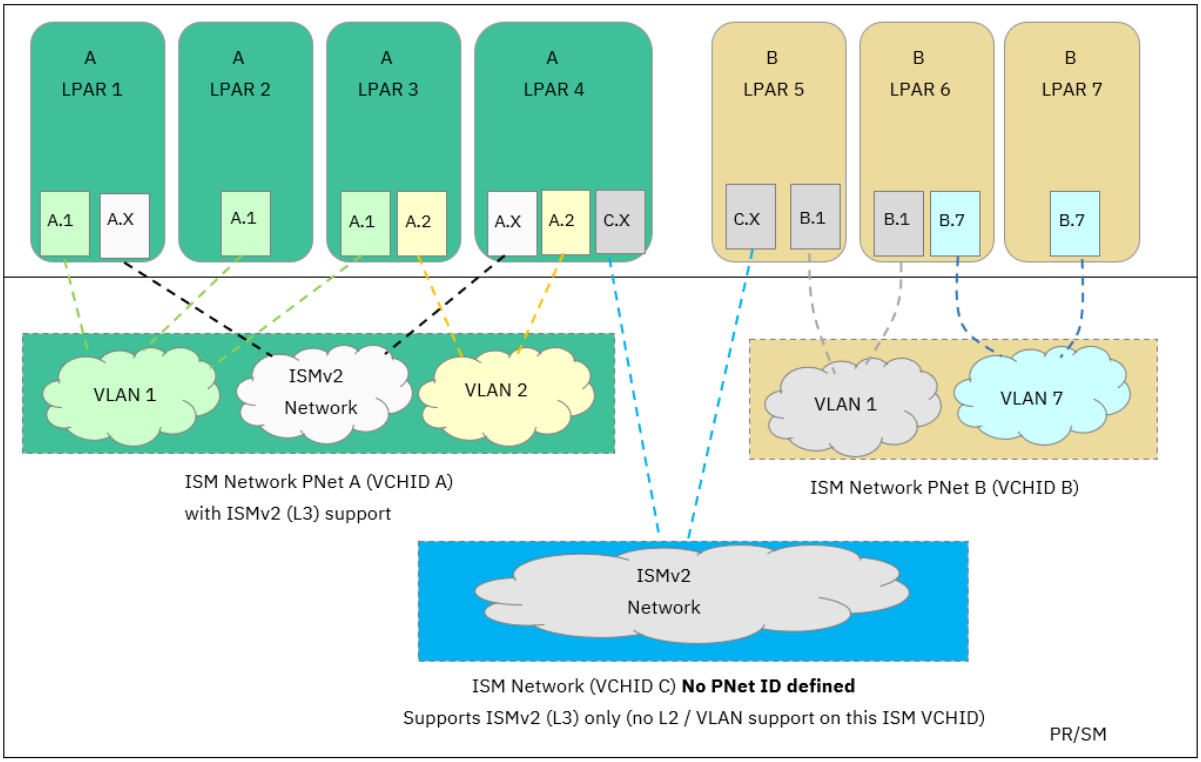


Figure 79. SMC-Dv2 and ISMv2 With and Without PNetIDs

While both (associated or non-associated) ISM configuration methods can be used to exploit SMC-Dv2 connectivity (enabling SMC-D over multiple IP subnets), the administrator must consider the entire set of systems that are required to communicate over the same ISM VCHID. When any system in this set

is down-level, method 1 (associated VCHIDs) is required (or at least combined as in method 3). While method 2 offers improvements in usability, it requires that all systems are up-level supporting SMC-Dv2.

SMC-Rv2

SMC-Rv2 defines the updates to the SMC-R protocol allowing SMC-Rv2 enabled hosts to connect and communicate over multiple IP subnets. SMC-Rv2 uses the RoCEv2 protocol also known as "routable RoCE". SMC-Rv2 connection eligibility is defined by the Enterprise ID. SMC-Rv2 peer hosts must define a common EID. SMC-Rv2 preserves many of the architectural concepts of SMC-R, such as supporting TCP sockets, dynamically creating SMC-Rv2 connections from the TCP/IP connection and preserving many of the related TCP quality of services such as the network administrative model, TCP connection load balancer compatibility, and TCP connection (TLS) security.

Some of the key differences of SMC-Rv2 are related to IP connectivity, such as provisioning an IP address to the RoCE interface and the reliance on your IP routing definitions (static or dynamic) and your IP topology (IP routes, routers, firewalls, etc.). In most cases your existing IP routing definitions will be sufficient. For high availability, it is strongly recommended that users define multiple equal cost routes between hosts exploiting SMC-Rv2.

Your RoCEv2 network high availability reuses your existing IP network high availability methodology. SMC-Rv2 continues to provide high availability through the SMC-Rv2 link group architecture using multiple links for load balancing and high availability.

SMC-R does not support IPSec. TCP connections that require IPSec are not eligible for SMC-Rv2 or SMC-Rv1. SMC-R network traffic supports connection level security such as TLS or AT-TLS.

RoCEv2

RoCEv2 defines the updates to the RoCE standards that enable "Routable RoCE". RoCEv2 encapsulates RoCE network traffic in UDP/IP packets using reserved UDP port 4791. RoCEv2 packets are not processed by z/OS.[6] All packet level processing is performed by the IBM RoCE Express2 feature. The RoCE Express2 feature on the IBM z15 provides the RoCEv2 support required by SMC-Rv2. RoCEv2 is supported on both the 10 and 25 GbE RoCE Express2 features.

The changes in SMC-Rv2 and RoCEv2 are transparent to z/VM. z/VM 7.1 or higher is required to support all versions of SMC-R.

Tip:

The z/OS administrator will not be required to reserve RoCEv2 UDP port 4791. When the administrator provisions a RoCEv2 (SMCR) IP address for the RoCEv2 interface in the OSA INTERFACE statement, the IP address is used exclusively by the RoCE Express hardware feature for RoCEv2 IP communications. Although the RoCEv2 IP address is provisioned and managed by the z/OS administrator, it is not used by z/OS. For this reason, the RoCEv2 IP address and UDP port will not conflict with a traditional z/OS IP address and port.

[6] z/OS can process UDP/IP RoCEv2 packets when the z/OS TCP/IP stack is acting as an IP router (datagram forwarding enabled) in the IP network path between RoCEv2 endpoints.

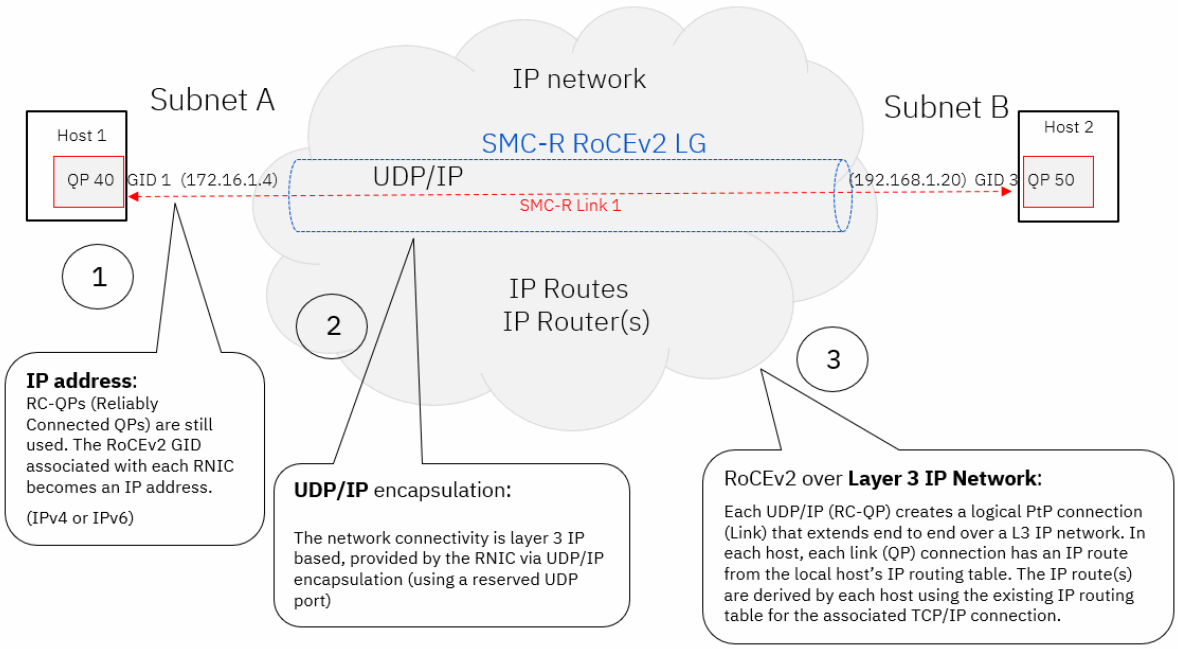


Figure 80. SMC-Rv2 connectivity concepts

Enabling SMC-Rv2

Enabling SMC-Rv2 involves the following two z/OS TCP/IP profile configuration steps:

1. Defining an Enterprise ID

   When the administrator defines EID in the TCP/IP profile on the GlobalConfig statement, it enables SMC-Rv2 for this TCP/IP stack for IPv4.[7]

2. Defining IP interfaces for RoCEv2

   The administrator must enable the applicable OSA (IPAQENET) interfaces for RoCEv2. When an EID is defined, enabling a single OSA interface can create SMC-Rv2 connections with SMC-Rv2 peer hosts. For high availability, it is recommended that at least two OSA interfaces are enabled for SMC-Rv2. See "SMC-Rv2 resilience, high availability with RoCEv2 and IP routing" on page 552 for more information.

   The required SMC-Rv2 definitions on the OSA INTERFACE[8] statement under the SMCR parameter are:

   a. PFID: PCI function ID that references the RoCE Express feature that supports RoCEv2 to be associated with this OSA IP interface. The PFID is defined in HCD.

   b. SMCRIPADDR: SMC-Rv2 IP address (RoCEv2 IP Address) that is to be used by the RoCE Express feature for RoCEv2 network traffic. This IP address must be the same IP version as this OSA and must be provisioned within the OSA IP subnet.

      Note that the SMCRMTU (RoCEv2 MTU) is an optional parameter.

      Tip:

      Careful consideration should be given to changing the default SMCRMTU size of 1 K. Increasing the SMCRMTU to a value greater than 1 K (e.g. 2 K or 4 K) requires that the entire IP network path, which can be multiple IP subnets, supports jumbo frames.

The above two steps (EID and INTERFACE) can be performed in either order. Once SMC-Rv2 is enabled in the stack with an EID and an eligible IP interface is active, then as new TCP/IP connections are created over this OSA IP interface, the TCP/IP connections become eligible for SMC-Rv2 when connecting with SMC-Rv2 enabled peer hosts sharing a common EID.

In order for a TCP/IP connection to be eligible for SMC-Rv2, the following conditions must all be true for the TCP client and the server, both hosts must:

- Be enabled for SMC-Rv2 (exchanged in SMC connection setup)
- Be configured with the same user-defined EID (exchanged in the SMC connection setup)
- Have IP routes defined over the associated OSA IP interfaces that are SMC-Rv2 enabled
- Not be prevented from SMC connectivity due to the presence of an SMC filter
- Not require IPSec for the associated TCP/IP connection

Tip:

Some firewalls can be configured to drop TCP options. The SMC protocol requires TCP option 254 to flow on the TCP/IP 3-way handshake. It is recommended that you verify that your firewalls permit TCP option 254 to flow. For more information about the SMC protocol, see RFC 7609.

Tip:

It is also recommended that you verify that UDP port 4791 (RoCEv2) is open in your firewalls.

OSA and RoCEv2 device association and physical connectivity

The OSA INTERFACE statement defines the specific physical OSA through the OSA interface port statement which references a VTAM TRLE. The SMCR RoCEv2 PFID parameter defines the specific RoCE feature. The OSA INTERFACE SMCR RoCEv2 parameters (PFID and SMCRIPADDR) logically bond the two adapters together. The TCP/IP communications is provided by OSA and the related SMC-Rv2 /

[7] SMC-Rv2 does not support IPv6.

[8] The OSA INTERFACE statement must meet all existing OSA INTERFACE definition requirements and all existing SMC-R definition requirements, such as defining a subnet mask. The associated OSA and RoCEv2 PFID must both have a matching PNetID (HCD).

RoCEv2 communications are provided by the RoCE adapter. The IP routes that will be reused for RoCEv2 connectivity will be associated with the OSA IP interfaces.

Both the OSA and associated RoCE features must:

- Have their physical ports defined with the same PNetID in HCD and
- Both physical ports must be attached on the same physical LAN (typically the same physical switch) and
- Both physical Ethernet switch ports must be defined with the same VLAN ID (when in trunk mode).

  The OSA INTERFACE VLAN ID definition is inherited by the RoCE PFID. When the OSA port is configured in access mode (VLAN ID is not known by the host), the RoCE port must also be configured in access mode.

  Rule: The RoCEv2 VLAN ID must match the VLAN ID of any adjacent (next hop) IP routers used by IP routes for RoCEv2 traffic. The RoCEv2 VLAN ID is inherited from the associated OSA interface and is either:

  – Zero, unknown by z/OS and is untagged (switch port is in access mode) or
  – A specific value (matching the switch port that is configured in trunk mode)

  Tip:

  To accommodate adapter sharing and various virtualization use cases, it is recommended that both the OSA and RoCEv2 ports are configured in trunk mode. SMC-Rv2 does support access mode, but it will limit your ability to share the physical adapters.

  Tip:

  Mixing 10 GbE and 25GbE RoCE adapters is supported on a LAN (direct connections), but you need to be aware of the potential for switch port congestion and bottlenecks forming.

  This mixture is also supported for multiple subnet networks. The potential for congestion exists for IP routers using 10 GbE adapters with hosts generating traffic at 25 GbE. Users should assure that the infrastructure associated with the entire IP route/path can accommodate the required bandwidth of all network traffic end to end over the entire IP route.

  Tip:

  SMC-Rv2 RoCEv2 traffic is encapsulated in UDP/IP by the "RoCE Express" feature. The "RoCE Express" feature is also responsible for managing the UDP network flow control by making adjustments to the UDP congestion management algorithms. The "RoCE Express" feature uses various network indications to adjust the network flow control algorithms. One such indication is defined by the Explicit Congestion Notification (ECN) standard.

  ECN is an existing industry standard congestion control mechanism applicable to the IP protocol. ECN provides notification (within the IP header) when IP packets are traversing an IP network path that is experiencing congestion. The notifications provided by ECN allow the network endpoints ("RoCE Express") to make adjustments to the flow control algorithms. The "RoCE Express" feature supports ECN by default. There is no action required on the z/OS host.

  It is recommended that SMC-Rv2 users enable ECN in their Ethernet switches and IP routers in the RoCEv2 network path defined by your IP routes.

  For more information about ECN, refer to RFCs 3168 and 8087. For information about the support of and the enabling of ECN, refer to the product support documentation provided with your network hardware switch or router equipment.

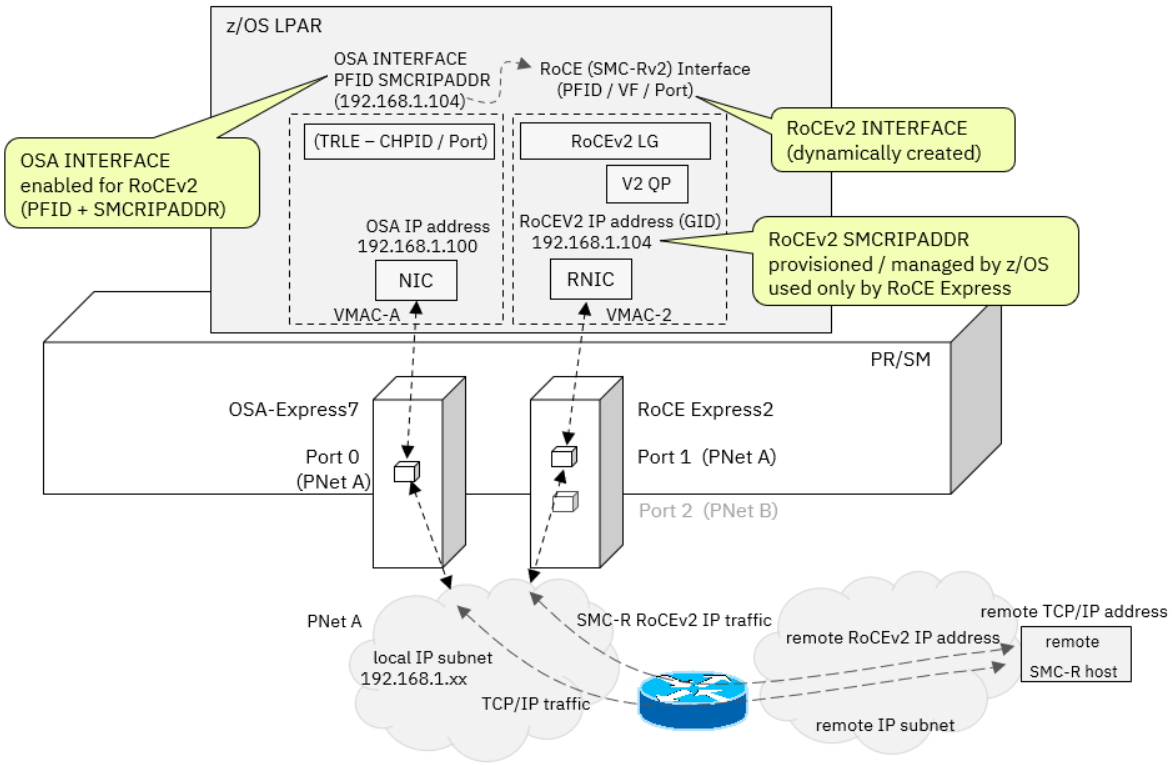


Figure 81. SMC-Rv2 OSA and RoCE relationship

SMC-Rv1 migration considerations

z/OS SMC-Rv2 users can control the use of SMC-Rv1 communications. Users who must continue to communicate with down-level SMC-R peer hosts which only support SMC-Rv1 (over the same IP subnet) must also configure the same FID value defined on OSA INTERFACE SMCR PFID parameter on the GlobalConfig SMCR statement to enable base SMC-R support. When the RoCE PFID is not configured on the GlobalConfig statement, SMC-Rv1 communications are disabled.

Result:

SMC-Rv2 enabled peers that are attached to the same IP subnet will use SMC-Rv2 / RoCEv2 on the same subnet.

Sharing or reusing resources: SMCRIPADDR and PFID

In some configurations, it is beneficial to reuse or share the SMCRIPADDR and the RoCEv2 PFID.

The SMCRIPADDR (RoCEv2 IP address) can be shared among multiple OSA interfaces associated with unique physical OSAs when the following conditions are true of all OSA INTERFACE statements:

1. Define the same subnet mask and
2. Define the same PFID value (reuse the same RoCEv2 adapter) and
3. Are associated with physical OSAs with ports attached to the same physical LAN (each OSA is defined with the same PNetID)

Rule:

The same RoCEv2 IP address cannot be shared among multiple OSA INTERFACE statements associated with the same physical OSA. In this case each INTERFACE requires a unique OSA IP address, subnet, and VLAN ID.

Rule:

The SMCRIPADDR (RoCEv2 IP address) is owned by a single z/OS TCP/IP stack. The same RoCEv2 IP address cannot be defined in (shared by) multiple TCP/IP stacks or used by any other LPARs.

The RoCEv2 FID can be shared among multiple OSA INTERFACE statements when the associated OSA ports are on the same physical LAN (each OSA is defined with the same PNetID). In the FID sharing case, the OSA INTERFACE statements can be associated with the same physical OSA or unique physical OSAs.

Tip:

A RoCE PFID allows unique users, typically LPARs, stacks or guests, to share (virtualize) the same physical adapter. There is no known reason or benefit for provisioning multiple PFID values associated with the same RoCE adapter and port to the same z/OS TCP/IP stack. This practice is not checked, prevented, or disallowed but it is not recommended.

SMC-Rv2 resilience, high availability with RoCEv2 and IP routing

SMC-Rv2 and RoCEv2 connectivity requires and is based on your IP routing definitions and your IP topology. Your IP routes can be statically or dynamically defined. Dynamic routing offers many significant advantages. SMC-Rv2 does not require unique IP routing design. Defining multiple equal cost IP routes is recommended. The SMC-Rv2 protocol reuses your existing TCP/IP connection IP routes. IP routes must be defined over one or more of your OSA IP interfaces.

SMC-Rv2 link groups

SMC-Rv2 preserves the network resiliency for workloads with the SMC-R link group architecture which has been adapted to RoCEv2. SMC-Rv2 LGs attempt to create two links (connections) over unique RoCEv2 adapters. When this can be achieved at both peers the LG is considered to be a symmetric LG. Symmetric LGs provide full redundancy when the connections are over unique adapters. The LG is created during the first TCP connection setup and it is reused for all subsequent TCP connections between the same

two hosts having the same TCP client and server roles and when both hosts use the same RoCEv2 IP addresses. The TCP server is responsible for managing the LG.

For the most part, your RoCEv2 network resiliency is achieved by reusing your existing IP network resiliency model. In order to achieve full SMC-Rv2 redundancy at each host and within the IP network, the following must be provided:

- Two OSA INTERFACE statements that are enabled for SMC-Rv2 (RoCEv2) with a PFID and SMCRIPADDR
- Each OSA INTERFACE statement is defined and associated with a unique physical RoCE adapter (PFID)
- Two or more distinct equal cost IP routes to the peer SMC-Rv2 hosts over the OSA IP interfaces and
- Each IP route is designed to have some form of hardware redundancy for your IP routing paths (e.g. redundant switches and IP routers).

Rule:

The IP routes used to create the SMC-R2 Link Groups come from the TCP/IP stack routing table. SMC-Rv2 does not use policy-based routes or the policy-based routing table to form v2 LGs.

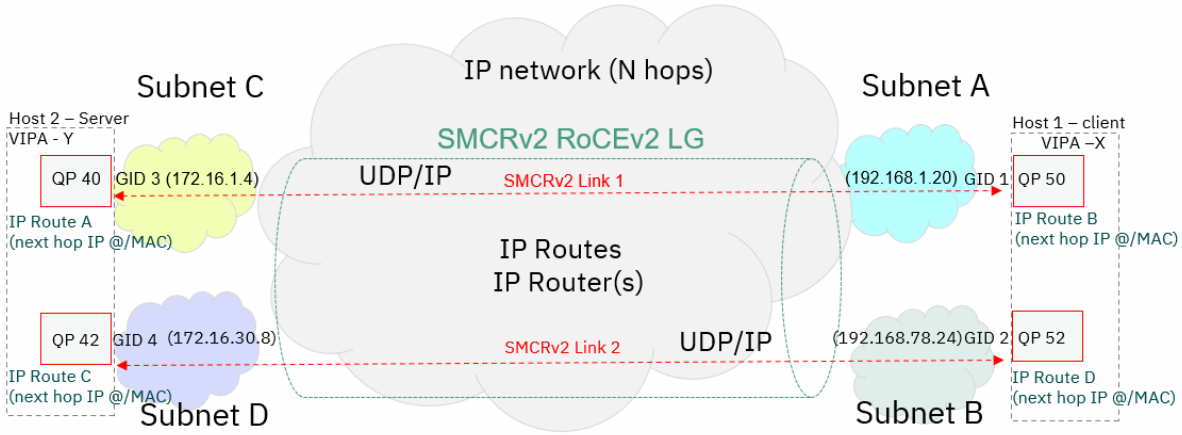


Figure 82. SMC-Rv2 link group

Although the OSA IP interfaces are not shown in Figure 82 on page 554, each OSA interface would be within the same IP subnets associated with each of the RoCEv2 links (QPs). This figure illustrates a

typical IP network configured for high availability. By reusing the IP network configuration, the SMC-Rv2 symmetric Link Group is dynamically formed to provide RoCEv2 high availability.

Tip:

The SMC-Rv2/RoCEv2 resiliency is dependent on both provisioning redundant hardware and your IP network configuration. TCP/IP NetStat link group display provides the level of redundancy for an SMC-Rv2 link group. The redundancy level is based on the host's local view of the RoCEv2 hardware (two SMC-Rv2 links having unique adapters, adapters are in unique I/O drawers and the LG type, such as symmetric versus asymmetric). See Netstat SMC-R LG report for more details.

Rule:

When only a single IP interface, a single IP route or a single RoCEv2 capable adapter is available, TCP/IP will create an SMC-Rv2 LG as either a single LG or an asymmetric LG. The TCP server will then periodically attempt to add a second link transitioning the LG to full redundancy (symmetric).

Rule:

When a direct route to a peer host can be found, it is preferred. When hosts are attached to the same IP subnet and both peers are SMC-Rv2 enabled, then SMC-Rv2 protocol is used. This is defined as an SMC-Rv2 direct link within the LG.

Similar to SMC-Rv1, SMC-Rv2 also provides for keep-alive support for both the TCP/IP connection and the RoCEv2 connection. In most use cases, the TCPMINKEEP setting on the GlobalConfig statement will be sufficient. For SMC-Rv2 indirect (multi-hop) links the TCP keep alive setting is set to 5 minutes, and for direct links the v1 TCPMINKEEP value is used.

Tip:

When SMC is used, the associated TCP connection stays active but idle. Some users exploit idle disconnect timers in their IP routers or firewalls. When SMC-Rv2 is used for multiple hop TCP/IP connections, the TCPMINKEEP setting should be set lower than your router disconnect timers.

SMC interactions with other z/OS Communications Server functions

SMC interacts with the following functions:

- "Sysplex distributor" on page 555
- "Security functions" on page 556
- "Intrusion detection services (IDS)" on page 556
- "TCP keepalive" on page 557
- "TCP application data transfer options" on page 558
- "Packet trace" on page 558
- "SMC-R RoCE maximum transmission unit" on page 558

Sysplex distributor

You can use Shared Memory Communications (SMC) with the sysplex distributor function with no additional configuration requirements. TCP connections are set up as normal through the sysplex distributor node, and the connection endpoints exchange SMC rendezvous information as described in "Rendezvous processing" on page 507. The sysplex distributor node examines information that is passed in the rendezvous exchange to select the optimum server application for SMC protocols. After the TCP connection is established to the server application, rendezvous processing continues to determine whether SMC communications are possible, and to establish an SMC link if necessary. After the TCP connection switches to SMC communications, the data does not flow through the sysplex distributor node, as shown in Figure 83 on page 556. This bypassing of the distributor node represents a performance improvement for the sysplex distributor function when SMC protocols are used.

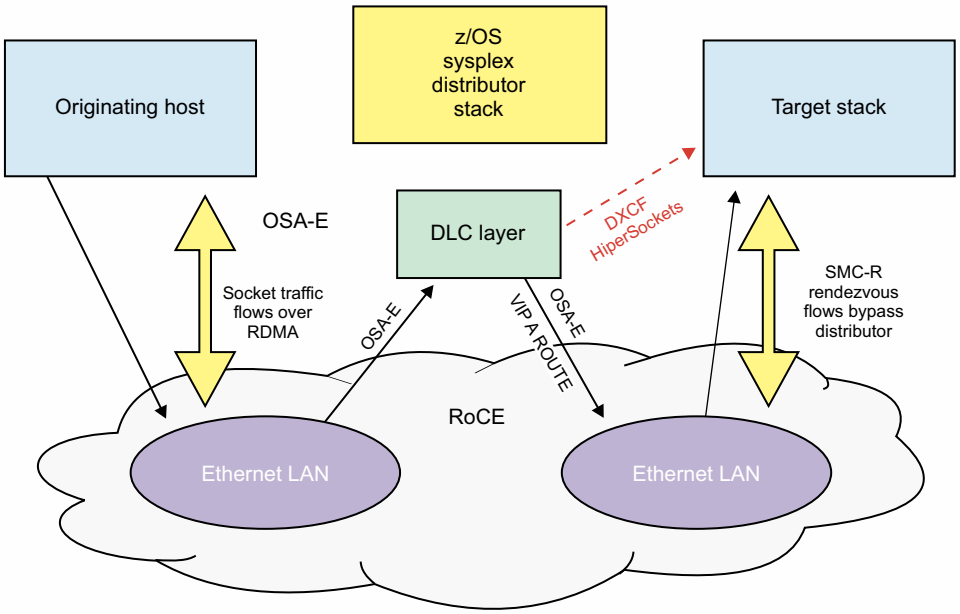


Figure 83. Sysplex distributor and SMC-R interaction

Security functions

You can use Application Transparent Transport Layer Security (AT-TLS) with SMC with no restriction. The negotiation of AT-TLS support for a TCP connection takes place after the SMC rendezvous exchange, and after that the encryption and decryption of the data occur as normal.

Generally, security functions that require TCP/IP to examine TCP packets cannot be used with SMC communications because data that is sent over SMC links is not converted into TCP packets. The following security functions do not interoperate with SMC:

- IPSec when the TCP connection requires tunneling
- IP filters when the filter would deny a packet for the TCP connection
- Multilevel security when the connection requires packet tagging (that is, when the source and destination zones are both SYSMULTI)

All of the above functions can be activated dynamically through profile changes, with the expectation that the change applies to current active TCP connections. If TCP connections are currently traversing SMC links when such a profile change is made, then the stack stops the TCP connections that are not compatible with the new security profiles.

Intrusion detection services (IDS)

Intrusion detection services (IDS) provides the following support:

- Scan detection and reporting
- Traffic regulation for TCP connections and UDP receive queues
- Attack detection, reporting, and prevention

The first two services involve checks that occur during TCP connection setup that do not interact directly with SMC communications.

The last service covers a range of checks. Most of the checks occur for inbound TCP packets, which means they are not applicable for SMC communications. The following two checks that are included in this service apply to TCP connections that traverse SMC links:

- TCP queue size events

You can use IDS policy to detect when the send or receive queue for a TCP connection that traverses an SMC link becomes constrained because of the amount or age of the data on the queue. When a queue becomes constrained, you can reset the TCP connection or continue to monitor the condition until the queue is no longer constrained.

Send or receive queues can be constrained for TCP connections that traverse SMC links:

- The send queue is considered to be constrained when data is available to be sent but cannot be sent, or when data is stored into the peer remote memory buffer (RMB) or direct memory buffer (DMB) that is not acknowledged for more than 30 seconds.
- The receive queue is considered to be constrained when data is available to be delivered but the application does not receive the data for more than 30 seconds.
- When either queue becomes constrained, the TCP connections are monitored or stopped based on the IDS policy in effect.

- Global TCP stall events

  You can use IDS to detect attacks that are designed to consume system resources by creating many TCP connections and causing them to stall, making them unable to send data. A global stall condition is in effect when at least 50% of the active TCP connections are stalled and at least 1000 TCP connections are active. You can reset stalled connections, or continue to monitor the condition.

  TCP connections that traverse SMC links are considered for global TCP stall events. A TCP connection that traverses an SMC link is treated as a stalled connection when the TCB is write-blocked.

TCP keepalive

Some intermediate nodes (for instance, load balancers or firewalls) use data traffic as an indication that the TCP connection is still alive. If no data flows across the TCP path for a long enough period, the intermediate node might reset the connection. TCP keepalive processing periodically sends packets over the TCP connection to prevent the connection from being reset.

The rendezvous negotiations to use SMC communications occur over the TCP connection. After the decision to switch to SMC protocols is made, the TCP connection remains active, but only termination messages flow over the TCP connection as needed. The application socket data flows out of band by using RMDA or ISM protocols. Thus, a TCP connection that is using SMC can be viewed as using two paths:

- A TCP path that is used for non-data packets, including the initial three-way handshake packets, FIN packets, and RST packets.
- An SMC path that is used for data.

Because the TCP connection does not use the TCP path to exchange data packets, intermediate nodes might consider the connection to be idle for long periods. For TCP connections over SMC, keepalive processing must ensure that both the TCP path and the SMC path remain operational.

For traditional TCP connections, the time interval that is used to send keepalive probes is determined by using the following sequence:

1. The value of the TCP_KEEPALIVE setsockopt() option, if specified by the application
2. The value of the INTERVAL parameter on the TCPCONFIG statement

For a TCP connection that is traversing an SMC link, the time interval that is used to send keepalive probes on the SMC path is determined by using the same method because the SMC path is where the actual data flows. For these connections, however, use of this same method for the TCP path can generate excessive keepalive probe traffic, so a separate method is used to determine the keepalive time interval for the TCP path.

- For TCP connections that use SMC-R communications, this method uses the larger of the GLOBALCONFIG SMCR TCPKEEPMININTERVAL value, the TCPCONFIG INTERVAL value, and the TCP_KEEPALIVE setsockopt() value.

- For TCP connections that use SMC-D communications, this method uses the larger of the GLOBALCONFIG SMCD TCPKEEPMININTERVAL value, the TCPCONFIG INTERVAL value, and the TCP_KEEPALIVE setsockopt() value.

For example:

- SO_KEEPALIVE is specified by the application, TCPCONFIG INTERVAL is 120 minutes, and GLOBALCONFIG SMCR TCPKEEPMININTERVAL value is 5 minutes.

  In this case, the time interval for both the SMC-R path and the TCP path is 120 minutes.
- SO_KEEPALIVE is specified by the application, TCPCONFIG INTERVAL is 10 minutes, TCP_KEEPALIVE setsockopt() is specified by the application with a value of 5 minutes, and the GLOBALCONFIG SMCD TCPKEEPMININTERVAL value is 15 minutes.

  In this case, the time interval for the SMC-D path is 5 minutes, but the time interval for the TCP path is 15 minutes.

TCP application data transfer options

The SMC protocol is transparent to and fully compatible with TCP socket applications. A wide range of functions are available through the API interfaces that z/OS Communications Server provides. The use of SMC protocols for exchanging application socket data is not apparent to the applications. Thus, socket API functions such as MSGWAITALL, MSG_PEEK, Accept and Receive (ANR), and Urgent Data can still be used by the applications, even when SMC communications are used.

Packet trace

Packet trace provides a mechanism for capturing the contents of TCP packets as they are sent and received. Even though SMC does not create TCP packets for data that is sent over SMC links, the data is captured as part of packet trace processing. This data includes all TCP connection data that is sent across the SMC link, and any SMC-R Link Layer Control (LLC) messages that the TCP/IP stack generates to manage the link or the associated memory buffers.

For information about formatting the SMC data in the packet trace, see z/OS Communications Server: IP Diagnosis Guide.

SMC-R RoCE maximum transmission unit

RDMA protocols, including SMC-R, define a unique set of supported maximum transmission unit (MTU) sizes. SMC-R protocols support MTU sizes of 256, 512, 1024, 2048, and 4096 bytes, but z/OS Communications Server requests an MTU size of at least 1024 bytes. When an SMC-R link is initially established between two peer hosts, the MTU size is exchanged and negotiated to the lowest value for both hosts. The negotiated MTU size must account for transport headers and cyclic redundancy check (CRC) information that is used by the underlying RoCE protocols.

Guidelines:

- The default MTU value for Communications Server is 1024, which can be used for most z/OS application workloads.
- In some cases, such as streaming or bulk data transfer, you can improve throughput by using an MTU setting of 2048 or 4096. You can configure the MTU value on the PFID subparameter of the GLOBALCONFIG statement.
- If you set the MTU size to 2048 or 4096, you must also enable jumbo frames on all switches in the path for all peer hosts.
- Define the same MTU size for all PFIDs that are associated with the same physical network (PNetID).

Managing SMC communications

You can use the VARY command to manage your "RoCE Express" and ISM interfaces, and the Netstat command to display Shared Memory Communications (SMC) information. You can use various tools to

monitor SMC information, and VTAM DISPLAY commands to display information about "RoCE Express" and ISM interfaces. You can stop SMC at a stack-wide level.

- “Managing your "RoCE Express" interfaces” on page 559
- “Managing your ISM interfaces” on page 560
- “Displaying SMC information” on page 560
- “Monitoring SMC information” on page 562
- “VTAM displays and tuning statistics” on page 563
- “Stopping SMC-R” on page 564
- “Stopping SMC-D” on page 565

Managing your "RoCE Express" interfaces

The TCP/IP stack dynamically creates "RoCE Express" interfaces for each configured PCI Express function ID (PFID). The stack activates all the configured "RoCE Express" interfaces when the first SMC-R capable IPAQENET or IPAQENET6 interface is started. After the "RoCE Express" interfaces are started, they remain active even when all SMC-R capable IPAQENET and IPAQENET6 interfaces are stopped. You can explicitly stop a "RoCE Express" interface by using the VARY TCPIP,,STOP command. The name of the dynamically created "RoCE Express" interfaces is in the form EZARIUT*pffff*, where *p* is the port number and *ffff* is the PFID.

Restriction: If you manually stop a "RoCE Express" interface, you must later restart that interface by using the VARY TCPIP,,START command. The stack does not automatically restart the "RoCE Express" interface when the next SMC-R capable interface is started.

In addition to stopping and starting the "RoCE Express" interface, you can dynamically add or delete PFIDs from your TCP/IP profile by using the VARY TCPIP,,OBEYFILE command.

Steps for dynamically adding a "RoCE Express" interface

Use these steps to dynamically add a "RoCE Express" interface.

Procedure

Perform the following steps:

1. Install and configure the "RoCE Express" feature in the hardware configuration definition (HCD).

   Configuration includes the physical network ID (PNetID) for the "RoCE Express" port. For more information, see “Physical network considerations” on page 520.
2. Add the new PCI Express function ID (PFID) to the GLOBALCONFIG SMCR parameter in the TCP/IP profile data set.

   Rule: In addition to the PFID that you are adding, you must also specify all currently configured PFIDs on the GLOBALCONFIG SMCR parameter. When you update the configured PFIDs on the GLOBALCONFIG SMCR parameter, you are completely replacing the configured PFID information.
3. Issue the VARY TCPIP,,OBEYFILE command and specify the updated TCP/IP profile data set.

Results

If any SMC-R capable IPAQENET and IPAQENET6 interfaces were ever active, the "RoCE Express" interface that the added PFID represents is automatically started. Otherwise, the interface is started when the first SMC-R capable interface is started.

What to do next

For more information about dynamically updating the GLOBALCONFIG SMCR parameter, see z/OS Communications Server: IP Configuration Reference.

Steps for dynamically removing a "RoCE Express" interface

Use these steps to dynamically remove a "RoCE Express" interface.

Procedure

Perform the following steps:

1. Issue the VARY TCPIP,,STOP command for the "RoCE Express" interface to be deleted.

   Result: If no alternative SMC-R link exists for connections that are using this "RoCE Express" interface, the connections are reset.
2. Remove the PCI Express function ID (PFID) value from the GLOBALCONFIG SMCR parameter in the TCP/IP profile data set.
3. Issue the VARY TCPIP,,OBEYFILE command and specify the updated TCP/IP profile data set.

What to do next

For more information about dynamically updating the GLOBALCONFIG SMCR parameter, see z/OS Communications Server: IP Configuration Reference.

Managing your ISM interfaces

The TCP/IP stack dynamically creates and activates an internal shared memory (ISM) interface for a specific physical network ID (PNetID) when both of the following conditions are true:

- The first SMC-D capable IPAQENET, IPAQENET6, IPAQIDIO, or IPAQIDIO6 interface is started for that PNetID.
- An ISM device with the same PNetID value is found.

After the ISM interface is started, it remains active even when all SMC-D capable interfaces are stopped. You can explicitly stop an ISM interface by using the VARY TCPIP,,STOP command. The name of the dynamically created associated ISM interfaces is in the form of EZAISM*xx*, where *xx* is a value from 01 to 28. Up to four unassociated ISM interfaces might be dynamically created and activated when the first SMC-D capable interface is activated.

Restriction: If you manually stop an ISM interface, you must later restart that interface by using the VARY TCPIP,,START command. The stack does not automatically restart the ISM interface when the next SMC-D capable interface is started. If you stop and then restart the ISM interface, z/OS Communications Server might assign a different Peripheral Component Interconnect Express (PCIe) function ID (PFID) to the ISM interface, depending on the availability of ISM devices that are associated with this PNetID.

Displaying SMC information

You can use the Netstat command and the DISPLAY TCPIP,,STOR command to display Shared Memory Communications (SMC) information.

For more information about these commands, see z/OS Communications Server: IP System Administrator's Commands.

Netstat ALL/-A report

You can use the Netstat ALL/-A report to determine whether a TCP connection is using SMC communications. You can filter the report by using the SMCID/-U filter and specifying an SMC-R link group ID, SMC-R link ID, or SMC-D link ID value as the filter. When a filter is specified, only those TCP connections that traverse the specified link group or link are displayed.

Netstat ALLConn/-a and Netstat COnn/-c reports

You can use the Netstat ALLConn/-a and Netstat COnn/-c reports to obtain TCP connection information. You can filter the reports by using the SMCID/-U filter and specifying an SMC-R link group ID, SMC-R link

ID, or SMC-D link ID value as the filter. When a filter is specified, only those TCP connections that traverse the specified link group or link are displayed.

Netstat CONFIG/-f report

You can use the Netstat CONFIG/-f report to display the settings on the GLOBALCONFIG statement, which includes the SMCR and SMCD parameters and subparameters.

Netstat DEvlinks/-d report

You can use the Netstat DEvlinks/-d report to display the following information about interfaces:

- Whether an IPAQENET or IPAQENET6 interface is eligible for SMC-R, and if so, information about the associated "RoCE Express" interfaces, also known as associated RNIC interfaces, that are in use.
- Whether an IPAQENET, IPAQENET6, IPAQIDIO, or IPAQIDIO6 interface is eligible for SMC-D, and if so, information about the associated internal shared memory (ISM) interface that is in use.

The physical network ID (PNetID) is also displayed, which can be useful in determining why "RoCE Express" or ISM interfaces are not associated with the OSA or HiperSockets interface.

You can also use the Netstat DEvlinks/-d report to display information about SMC links and link groups:

- You can display the SMC-R links and SMC-R link groups that traverse a "RoCE Express" interface, including the redundancy level for the link group as described in “SMC-R high availability considerations” on page 522.
- You can display the SMC-D links that traverse an ISM link.

To obtain information about SMC links and link groups, you must specify the SMC modifier on the Netstat command.

You can filter the Netstat report by using the SMCID/-U filter and specifying an SMC-R link ID or link group ID as the filter:

- When an SMC-R link ID is specified, information about that SMC-R link is displayed, in addition to information about the SMC-R link group that contains the link.
- When an SMC-R link group ID is specified, information about the SMC-R link group and all SMC-R links within the group is displayed.
- When an SMC-D link ID is specified, information about that SMC-D link is displayed.

You can display information about interfaces based on the PNetID value that is assigned to the interface by using the PNETID modifier on the Netstat command.

- To display all interfaces that have a PNetID value, specify PNETID=* as the modifier value.
- To display all interfaces that have a specific PNetID value, specify PNETID=*physical network ID* as the modifier value.

For more information, see z/OS Communications Server: IP System Administrator's Commands.

Netstat PORTList/-o report

You can use the Netstat PORTList/-o report to display the settings of the NOSMC and SMC parameters for the defined TCP port or range of ports.

Netstat STATS/-S report

You can use the Netstat STATS/-S report to display SMC-R and SMC-D usage statistics.

- In some cases, the statistics are a subset of the overall TCP statistics of the same name. For example, a value for the field `Current Established Connections` is displayed under `TCP Statistics`, `SMCR Statistics`, and `SMCD Statistics`. To determine the number of current connections that do not traverse an SMC link, subtract the total of SMC-R and SMC-D statistics from the total of TCP statistics.

- In other cases, the statistics represent information that is applicable to SMC-R or SMC-D processing only. For instance, the field `Active SMC Links Opened` has no corresponding TCP statistics value.
- For information about which SMC statistics fit in which category in the Netstat STATS/-S report, see z/OS Communications Server: IP System Administrator's Commands.

DISPLAY TCPIP,,STOR output

You can use the DISPLAY TCPIP,,STOR command to display current SMC-R and SMC-D storage usage.

- SMC-R values are displayed for remote memory buffers (RMBs), denoted as `SMC-R RECV MEMORY`, and staging buffers, denoted as `SMC-R SEND MEMORY`. This storage is limited by the setting of the FIXEDMEMORY subparameter on the GLOBALCONFIG SMCR parameter. The SMC-R information is displayed only when the SMC-R function is enabled or was previously enabled for this TCP/IP stack.
- SMC-D values are displayed for direct memory buffers (DMBs), denoted as `SMC-D FIXEDMEMORY`. This storage is limited by the setting of the FIXEDMEMORY subparameter on the GLOBALCONFIG SMCD parameter. The SMC-D information is displayed only when the SMC-D function is enabled or was previously enabled for this TCP/IP stack.

Monitoring SMC information

z/OS Communications Server provides various tools that you can use to monitor Shared Memory Communications (SMC) information:

- “Network Management Interface” on page 562
- “SMF records” on page 563
- “SNMP” on page 563

Network Management Interface

Several TCP/IP callable NMI (EZBNMIFR) requests include information about SMC in general and specifics about "RoCE Express" and ISM interfaces:

- The GetConnectionDetail request provides information about active TCP connections, including SMC information for TCP connections that traverse SMC-R or SMC-D links.
- The GetGlobalStats request provides TCP/IP stack global statistics for IP, ICMP, TCP, UDP, SMC-R, and SMC-D processing.
- The GetIfs request provides TCP/IP stack interface attributes and IP addresses, including information about "RoCE Express" and ISM interfaces.
- The GetProfile request provides profile information, including GLOBALCONFIG information for SMC-R and SMC-D.

In addition, the following NMI requests that are specific to SMC-R information are available:

- The GetRnics request, which provides interface statistics and VTAM tuning statistics that are related to "RoCE Express" interfaces. This request provides similar information to what is returned on the GetIfStats and GetIfExtendedStats requests for traditional interfaces.
- The GetSmcLinks request, which provides statistics about active SMC-R link groups and the SMC-R links within the link groups.

In addition, the following NMI requests that are specific to SMC-D information are available:

- The GetIsms request, which provides interface statistics that are related to ISM interfaces. This request provides similar information to what is returned on the GetIfStats and GetIfExtendedStats requests for traditional interfaces.
- The GetIsmLinks request, which provides statistics about active SMC-D links.

For specifics about these TCP/IP callable NMI requests, see z/OS Communications Server: IP Programmer's Guide and Reference.

SMF records

Several SMF 119 records include information about SMC in general and specifics about "RoCE Express" interfaces or ISM interfaces:

- The TCP connection termination SMF record (subtype 2) notifies the user when a TCP connection terminates. The record includes SMC information for TCP connections that traverse SMC-R or SMC-D links.
- The TCP profile event SMF record (subtype 4) notifies the user of changes to the TCP/IP profile. The records include GLOBALCONFIG and SMFCONFIG information for SMC-R and SMC-D, and IPCONFIG and IPCONFIG6 information for SMC-D.
- The TCP/IP statistics SMF record (subtype 5) provides statistics for a variety of TCP/IP functions. The records include global statistics and storage usage statistics that are related to SMC-R and SMC-D processing.

In addition, SMF type 119 records that are specific to SMC-R processing are available:

- The SMC-R link state start (subtype 42) and SMC-R link state end (subtype 43) SMF records notify the user when an SMC-R link is activated or deactivated. These SMF records serve a similar purpose for SMC-R links as the TCP connection start and TCP connection end records serve for TCP connections.
- The SMC-R link group statistics SMF record (subtype 41) provides statistics for all active SMC-R link groups on an interval basis. This SMF record includes information pertinent to both SMC-R link groups and the individual SMC-R links that comprise the link group.
- The RNIC interface statistics SMF record (subtype 44) provides interface statistics for "RoCE Express" interfaces on an interval basis. This SMF record serves the same purpose for "RoCE Express" interfaces that the interface statistics SMF record (subtype 6) serves for other interfaces.

In addition, SMF type 119 records that are specific to SMC-D processing are available:

- The SMC-D link state start (subtype 39) and SMC-D link state end (subtype 40) SMF records notify the user when an SMC-D link is activated or deactivated. These SMF records serve a similar purpose for SMC-D links as the TCP connection start and TCP connection end records serve for TCP connections.
- The SMC-D link group statistics SMF record (subtype 38) provides statistics for all active SMC-D link on an interval basis.
- The ISM interface statistics SMF record (subtype 45) provides interface statistics for ISM interfaces on an interval basis. This SMF record serves the same purpose for ISM interfaces that the interface statistics SMF record (subtype 6) serves for other interfaces.

For more information about type 119 SMF records, see z/OS Communications Server: IP Programmer's Guide and Reference.

SNMP

The Simple Network Management Protocol (SNMP) TCP/IP subagent provides Shared Memory Communications information, such as SMC eligibility and the associated physical network ID (PNetID) for IPAQENET, IPAQENET6, IPAQIDIO, IPAQIDIO6, RNIC, and ISM interfaces. SNMP also reports basic information about "RoCE Express" and ISM interfaces. The following SNMP MIB tables contain the information:

- ibmTcpipMvsIfTable
- ibmTcpipMvsPortTable

All the MIB objects that are supported by Communications Server functions are listed in the MIB objects appendix in z/OS Communications Server: IP System Administrator's Commands.

VTAM displays and tuning statistics

When a "RoCE Express" or internal shared memory (ISM) interface is first started, VTAM dynamically creates a transport resource list element (TRLE) to represent it.

- For a RoCE Express2 interface, the TRLE name is in the form of IUT*ppfff*, where *p* is the port number and *ffff* is the PCI-Express function ID (PFID).
  - For a 10 GbE RoCE Express feature, if you specify `GLOBALCONFIG SMCR PFID 0018 PORTNUM 1`, the TRLE name is IUT10018.
  - For a RoCE Express2 feature, if you specify `GLOBALCONFIG SMCR PFID 0055`, the TRLE name will be IUT10055 or IUT20055. It depends on the port number defined for PFID 55 in the Hardware Configuration Definition (HCD). VTAM learns the port number when the PFID is activated.
- For an ISM interface, the TRLE name is in the form of IUT0*ffff*, where *ffff* is the PFID that is associated with the ISM device that VTAM selects to use with this physical network. For example, if the PFID value is 0024, the TRLE name is IUT00024.

You can use the VTAM DISPLAY TRL and DISPLAY ID commands to display information about the TRLE representation of the "RoCE Express" or ISM interface, including information about the physical network ID (PNetID), and which TCP stacks use the interface.

Tip: For a 10 GbE RoCE Express TRLE, use the presence of a virtual function number (VFN) in the DISPLAY TRL or DISPLAY ID command output to determine whether the "RoCE Express" feature operates in a shared RoCE environment. A VFN is present in a shared environment and absent in a dedicated environment. A RoCE Express2 TRLE always has a VFN and always operates in a shared environment.

VTAM tuning statistics are managed differently for "RoCE Express" interfaces and ISM interfaces:

- VTAM collects tuning statistics for "RoCE Express" interfaces when requested by using the TNSTAT start option or the MODIFY TNSTAT command. Tuning statistics that represent processing at a "RoCE Express" interface level and statistics at a user or TCP/IP stack level are both maintained. The TCP/IP stack level statistics are also provided on the GetRnics request. For more information, see “Network Management Interface” on page 562.
- ISM interfaces are not affected by the MODIFY TNSTAT command or the TNSTAT start option. VTAM collects a minimal set of tuning statistics for the ISM interface, and these statistics are included on the Netstat DEvlinks/-d report and the GetIsms request. For more information, see “Network Management Interface” on page 562.

For more information about the VTAM commands, see z/OS Communications Server: SNA Operation. For more information about gathering tuning statistics, see z/OS Communications Server: SNA Network Implementation Guide.

Stopping SMC-R

You can stop Shared Memory Communications over RDMA (SMC-R) at a stack-wide level.

Procedure

Perform the following steps to stop SMC-R:

1. Modify the GLOBALCONFIG statement in the TCP/IP profile data set to include the NOSMCR parameter.
2. Issue the VARY TCPIP,,OBEYFILE,*profile_data_set* command.

Results

The TCP/IP stack does not immediately stop all existing SMC-R processing after the VARY TCPIP,,OBEYFILE command is processed. Existing TCP connections that are using existing SMC-R links are unaffected and continue to use SMC-R communications until they end or are stopped. However, no new SMC-R link groups or links are created, and no TCP connections that are established after this point use the existing SMC-R links.

What to do next

When you are restarting SMC-R after a stop, the previous SMCR parameter settings are used when they are not explicitly configured in the OBEYFILE data set. For more information about dynamically updating the GLOBALCONFIG SMCR parameter, see z/OS Communications Server: IP Configuration Reference.

Stopping SMC-D

You can stop Shared Memory Communications - Direct Memory Access (SMC-D) at a stack-wide level.

Procedure

1. Modify the GLOBALCONFIG statement in the TCP/IP profile data set to include the NOSMCD parameter.
2. Issue the VARY TCPIP,,OBEYFILE,*profile_data_set* command.

Results

The TCP/IP stack does not immediately stop all existing SMC-D processing after the VARY TCPIP,,OBEYFILE command is processed. Existing TCP connections that are using existing SMC-D links are unaffected and continue to use SMC-D communications until SMC-D links end or are stopped. However, no new SMC-D links are created, and TCP connections that are established after this point do not use the existing SMC-D links.

What to do next

When you restart SMC-D after a stop, the previous SMCD parameter settings are used if they are not explicitly configured in the OBEYFILE data set. For more information about dynamically updating the GLOBALCONFIG SMCD parameter, see z/OS Communications Server: IP Configuration Reference.

Part 2. Server applications

© Copyright IBM Corp. 2000, 2022

Chapter 11. Accessing remote hosts using Telnet

Telnet is a terminal emulation protocol. With Telnet, users can log on to remote host applications as though they were directly attached to that host. Telnet protocol requires that the user have a Telnet client that emulates a type of terminal that the host application can understand. The client connects to a Telnet server, which communicates with the host application. The Telnet server acts as an interface between the client and host application. A PC can support several clients simultaneously, each with its own connection to any Telnet server. This topic describes how to set up and use the following kinds of Telnet servers:

- TN3270E Telnet server

  Provides access to z/OS VTAM SNA applications on the MVS host using Telnet TN3270E, TN3270, or linemode protocol
- z/OS UNIX Telnet server

  Provides access to z/OS UNIX shell applications on the MVS host using Telnet linemode protocol

You can use the same port for both Telnet servers. For an overview of port management, see "Port management overview" on page 46. For more specific information about the PORT BIND statement, see "Setting up reserved port number definitions in PROFILE.TCPIP" on page 259.

The TN3270E Telnet server

The TN3270E Telnet server (Telnet) provides access to z/OS VTAM SNA applications on the MVS host using Telnet TN3270E, TN3270, or linemode protocol. Telnet acts as an interface between IP and SNA networks. End users in an IP network connect to Telnet, which is also a VTAM application. Telnet activates one SNA application minor node logical unit (LU) to represent each Telnet IP client. The Telnet application LU establishes a session with a VTAM host application (for example, CICS or TSO), simulating a terminal (LU0 or LU2) or a printer (LU1 or LU3). To enable connections, you must modify the VTAM and Telnet configuration data sets with Telnet statements. These statements describe Telnet server characteristics, the Telnet LUs, a listening port, and the characteristics of that port. After Telnet is started, you can use VARY and DISPLAY commands that are specifically related to Telnet to alter the state of Telnet or to display information about Telnet. For more information about these command sets, see z/OS Communications Server: IP System Administrator's Commands.

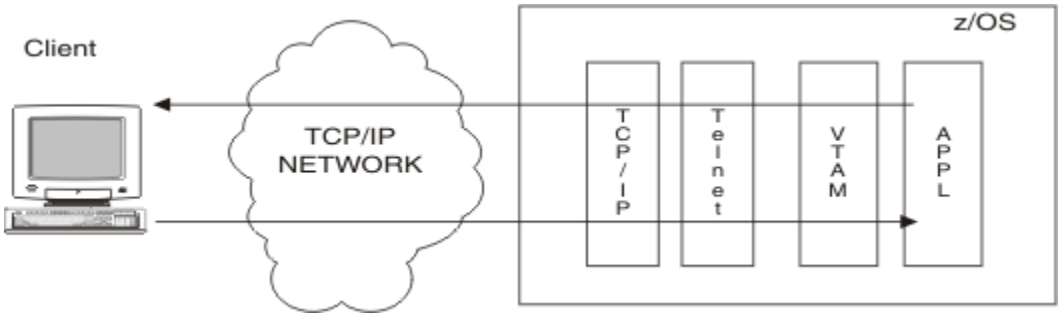


Figure 84. Telnet connectivity

Steps for starting the TN3270E Telnet server

Use these steps for the minimum information that you need to start the TN3270E Telnet server. The steps refer to more specific information about customizing your server.

Before you begin

You need to know how to create VTAM definition data sets; specifically, you need to know how to define VTAM application LUs. You need to set up and know how to use a TN3270E client emulator. A maximum of eight Telnet servers can be active at any time.

© Copyright IBM Corp. 2000, 2022

Procedure

Perform the following steps to start the TN3270E Telnet server:

1. Create a new data set member in your procedure library for the TN3270E Telnet server JCL.

   A sample procedure is in SEZAINST(EZBTNPRC). The only valid parameter that can be passed in from the JCL (using the PARM= keyword on the EXEC JCL statement) is the component trace options parmlib member name.
2. Define security for a user ID and associate the user ID with the Telnet procedure name; see "Steps for defining security for a user ID and associating the user ID with the Telnet procedure name" on page 571.
3. Customize the VTAM configuration data set to define VTAM application LUs for Telnet to use; see "Steps for customizing the VTAM configuration data set for Telnet" on page 573.
4. Customize the TCP/IP configuration data set.

   Reserve your Telnet ports by using the PORT *num tnproc* statement in the TCP/IP startup profile. If you do not code the PORT *num tnproc* statement, another application might use the port before the Telnet application can claim it.
5. Customize the TN3270E Telnet server configuration data set; see "The TN3270E Telnet server configuration data set" on page 574 and "Steps for customizing the TN3270E Telnet server configuration data set" on page 575.
6. Set the component trace options (CTRACE).

   Component trace options are set in a separate parmlib data set member. The sample procedure JCL points to a sample parmlib member, CTIEZBTN, in SYS1.PARMLIB, which starts a minimum trace. Use this member unless you need to turn on other options to debug a Telnet problem.

   To change the component trace options, specify a new parmlib member in the JCL. The member has the form CTIEZB*xx*. For more information about setting up trace options, see "Telnet CTRACE" on page 577.
7. Set up the resolver input file.

   Use the default search order unless there are special circumstances that require you to use unique parameters. For example, if there are parameters that are to be used only when the resolver is called by Telnet, you need to define those unique parameters. Define the unique parameters in a data set that is specified on the SYSTCPD DD statement in the Telnet procedure JCL. For more information about resolvers and resolver configuration files, see Chapter 13, "The resolver," on page 729.
8. Start Telnet by issuing START *tnproc*, where *tnproc* is the Telnet procedure member name.

Results

You should now see message EZZ6003I *tnproc* LISTENING ON PORT *nnnnn*. To verify your configuration, connect to the Telnet port from your Telnet client emulator; Telnet uses the first Telnet LU that is available. A solicitor screen or USSMSG10 screen prompts you to enter an application. Enter the name of any valid SNA application to log on to that SNA application. If an error occurs during session initiation, the MSG07 statement causes an error message to be sent to the client. If you do not code the MSG07 statement, then the connection is dropped.

Tips:

- To help avoid any unnecessary IPL, specify REUSASID=YES on the START command to ensure that the address space identifier (ASID) associated with the Telnet address space can be reused. The Telnet address space provides PC-entered services that must be accessible to all address spaces, so a system linkage index (LX) is obtained. Unless you specify REUSASID=YES on the START command, the ASID associated with the Telnet address space will be nonreusable when the address space is stopped or restarted. If the Telnet address space is stopped enough times and REUSASID=YES is not specified when started, all available ASIDs could be exhausted, preventing the creation of a new address space on the system. In this case, an IPL is required. For more information about tuning parameters for the maximum number of ASIDs on a system, see the MAXUSER parameter in z/OS MVS Initialization and Tuning Reference.

- The default MVS program properties table (PPT) entry sets Telnet to be a non-cancelable application. As a non-cancelable application, a TCP/IP stack should not automatically start Telnet using the AUTOLOG function. If the TCP/IP stack is recycled, the stack tries to cancel and restart all AUTOLOG applications. A non-cancelable application does not end and the following messages are issued repeatedly:

```
N 0140000 SA6I  2005147 04:59:27.69 STC07087 00000084  EZZ0621I AUTOLOG FORCING IBMTNSI0, REASON: TCP/IP HAS BEEN RESTARTED
NR0000000 SA6I  2005147 04:59:27.71 STC07087 00000080  IEE838I IBMTNSI0          NON-CANCELABLE - ISSUE FORCE ARM
```

  If you want to set the priority for Telnet in the PPT, change the priorities by assigning the job name to another service class in the STC subsystem.

  The default settings are: Privileged, non-swappable, non-cancelable, running in key 6, and system task. With these settings, Telnet and the TCP/IP stack have the same priority. The privileged or system task setting causes the started job to be assigned to the SYSSTC service class.
- To change the IPv6 or INET environments, you must recycle the Telnet procedure. Telnet checks for changes in the environment each time a port is activated. If Telnet detects a change in the environment, then the port is not activated. If you change from an IPv4 environment to an IPv6 environment, from an IPv6 environment to an IPv4 environment, from an INET environment to a CINET environment, or from a CINET environment to an INET environment, then results are unpredictable on existing ports.

Steps for defining security for a user ID and associating the user ID with the Telnet procedure name

Before you can start Telnet, you must define security for a user ID and associate it with the Telnet procedure name. These steps use RACF as the example security subsystem. If you are using another security product, see the documentation for that product to determine the appropriate procedure.

Before you begin

- You need to know the name of the Telnet procedure that you are using.
- Ensure that the MAXSOCKETS value is large enough to support the anticipated number of sockets that will be used by the system. If your system is IPv6 enabled, Telnet listening sockets are IPv6; set the IPv6 MAXSOCKETS value appropriately.
- Ensure that the file descriptor limit for a process is large enough to support the anticipated number of sockets that will be used by Telnet. If your system MAXFILEPROC limit is adequate, you do not need to make a change. If you need a larger limit, you can configure a higher limit for Telnet using the FILEPROCMAX attribute in the OMVS SAF options for the Telnet user ID. If the Telnet user ID is a superuser ID or has READ access to the BPX.SUPERUSER resource, Telnet will dynamically increase its file descriptor limit to the maximum allowed by z/OS UNIX System Services, currently 524,287. For more information about MAXFILEPROC, see z/OS UNIX System Services Planning.

Procedure

Perform the following steps to define security for a user ID and to associate it with the procedure name.

1. Use an existing user ID or create a new user ID:
   - Define a user ID with a nonzero UID value and do not permit it to the BPX.SUPERUSER resource. You will see message EZZ6049I *tnproc* NON-ZERO OMVS UID IN EFFECT, indicating that you associated Telnet to a user ID that does not have superuser authority. The number of connections allowed on a single port will be the MAXFILEPROC value. The number of connections can be overridden on the ALTUSER command with the FILEPROCMAX option. For example, you can allow 150,000 connections using the following command:

     ```
     ALTUSER TN3270E DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(23) FILEPROCMAX(150000)
     HOME('/'))
     ```

     If connection failures occur (EZZ6012I *tnproc* BPX1AIO ACCEPT FAILED, RC = 0000007C RSN = 050B0146) followed by a port quiesce (EZZ6003I *tnproc* QUIESCED ON PORT 23), the MAXFILEPROC value has been reached.

Tip: If your MAXFILEPROC value is less than your expected number of Telnet connections on a single port, you should use superuser authority or the FILEPROCMAX option on the RACF ALTUSER command. The FILEPROCMAX value will override the MAXFILEPROC value for processes associated with the user ID. If you do not use the FILEPROCMAX option and you do not give the associated user ID superuser authority by permitting the user ID to the BPX.SUPERUSER resource, Telnet is not able to increase the MAXFILEPROC value on the listener socket and will support the number of connections specified by the MAXFILEPROC value instead of the OMVS maximum.

- Permit a user ID with a nonzero UID value to the BPX.SUPERUSER resource in the FACILITY class:

  a. Add the user to RACF:

  ```
  ADDUSER TN3270E
  ```

  b. Permit the user ID:

  i) Create a BPX.SUPERUSER FACILITY class profile:

  ```
  RDEFINE FACILITY BPX.SUPERUSER
  ```

  ii) If this is the first class profile, activate the FACILITY class:

  ```
  SETROPTS CLASSACT(FACILITY)
  SETROPTS RACLIST(FACILITY)
  ```

  iii) Permit the user to the class:

  ```
  ALTUSER TN3270E DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(23) HOME('/'))
  PERMIT BPX.SUPERUSER CLASS(FACILITY) ID(TN3270E) ACCESS(READ)
  ```

  In this example, the user ID is TN3270E and the UID is 23. The UID can be any nonzero number. UID 23 was used to match the well-known Telnet port number.

  iv) Refresh the FACILITY class:

  ```
  SETROPTS RACLIST(FACILITY) REFRESH
  ```

  This example uses TN3270E for the user ID, but you can use any name.

  Tip: You can combine the ADDUSER and ALTUSER commands into one command by putting the OMVS parameter on the ADDUSER command line. The ADDUSER and ALTUSER commands are performed separately in case the user ID already exists. Even if the ADDUSER command fails, the ALTUSER command is successful.

- Use an existing superuser ID to associate with the job name.
- Define a superuser ID to associate with the job name.

  To define a superuser ID, add a user ID to RACF and alter it to superuser status:

  ```
  ADDUSER TN3270E
  ALTUSER TN3270E DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(0) HOME('/'))
  ```

  Sample statements for defining a superuser ID are in SEZAINST(EZARACF). For more information, see z/OS UNIX System Services Planning, z/OS Security Server RACF Security Administrator's Guide, and z/OS Security Server RACF Command Language Reference.

2. Add the procedure name to the RACF STARTED class and associate the user ID from step 1 with the name.

   For example, code the following statements:

   ```
   RDEFINE STARTED TELNET*.*  STDATA(USER(TN3270E))
   SETROPTS RACLIST(STARTED) REFRESH
   ```

   Sample statements for adding the procedure to the STARTED class are in SEZAINST(EZARACF). For more information, see z/OS UNIX System Services Planning, z/OS Security Server RACF Security Administrator's Guide, and z/OS Security Server RACF Command Language Reference.

3. If you are using secure Telnet connections, make sure that the user ID that runs Telnet has access to the SSL key ring and certificates.

   Take one of the following actions:

   - Create a new key ring and certificate that is owned by the user ID that runs Telnet.
   - Share the existing key ring and certificates with Telnet. For more information about implementation scenarios, see z/OS Security Server RACF Security Administrator's Guide.
4. If you are using hardware encryption, ensure that the Telnet user ID has read access to the RACF CSFSERV class resources. For details, see “Encryption algorithms” on page 1332.

Results

You know you are done when you can start Telnet without receiving errors.

If your job ends abnormally with system completion code EC6 and a register 15 value of 0F01C008, you did not associate a valid user ID with the started job name.

Steps for customizing the VTAM configuration data set for Telnet

Telnet uses application LUs that are defined in VTAM application (APPL) major nodes to represent clients.

Before you begin

You can code LU definitions in the data set members of your choice. You must include the data set that contains your member updates in the list of data sets that is specified on the VTAMLST DD statement in the procedure that is used to start VTAM. Including this data set in the list ensures that the LUs are available for activation after VTAM is started.

Procedure

Perform the following steps to customize the VTAM configuration data set:

1. Automatically activate the application definition deck by including it in ATCCONxx.
2. Define the LUs in a VTAM data set member.

   A sample VTAM configuration data set is in SEZAINST(IVPLU).

   Tip: In the VTAM configuration data set, you can define Telnet LUs that represent either terminal or printer emulators with a wildcard character instead of coding individual Telnet application LU statements. The Model Application Names function enables you to code a model APPL name with an asterisk (*) or a question mark (?) (see z/OS Communications Server: SNA Resource Definition Reference for more information). Use * as a wildcard character to replace a character string. For example, if you need Telnet LUs in the range TCPABC01-TCPABC99, then you can add a single VTAM application definition statement to the sample configuration data set. The definition statement has the Telnet application minor node (Telnet LU) name TCPABC*, which supports all 99 LUs.
3. Use the default value YES for the SESSLIM parameter.

   Telnet server LUs do not support multiple concurrent sessions.
4. Code LOSTERM=IMMED on all target (PLU) applications that will have a SNA session with Telnet.

   If you do not code this value, CLOSEACB might stop while it is waiting for UNBIND RESPONSE if the target (PLU) application does not issue CLSDST when the LOSTERM exit is processed.
5. Code EAS=1 to minimize Common Service Area (CSA) storage use.

   If you use the default value for EAS, excessive CSA storage will be used.
6. Use the default VTAM value NO for PARSESS.

   You should not use parallel sessions with Telnet LUs.

Results

When you are done, start VTAM. Put the LUs in a connectable state by activating the APPL major node that represents the Telnet LUs. Issue the D NET,MAJNODES command or the D NET,ID=*appl_major_node_name* command to verify that your APPL major node is active.

The TN3270E Telnet server configuration data set

Telnet configuration statements are processed during the initialization of the TN3270E Telnet server or when you issue the VARY TCPIP,*tnproc*,OBEYFILE command to update the Telnet configuration data set. (For information about using the VARY TCPIP,*tnproc*,OBEYFILE command to update the Telnet configuration data set, see "Using the VARY TCPIP,tnproc,OBEYFILE command to update Telnet configuration" on page 579). The Telnet configuration statements enable the following definition and session setup:

- Define Telnet server characteristics
- Define connection characteristics
- Define LU names to represent Telnet clients
- Facilitate session setup with MVS host VTAM applications

Requirement: When configuration data sets are processed, the TELNETPARMS and BEGINVTAM blocks are required for each port that you start or modify.

You use the following statement blocks to configure Telnet:

TELNETGLOBALS and ENDTELNETGLOBALS
: An optional statement block that contains Telnet parameter statements. The parameters define connection characteristics for all ports.

TELNETPARMS and ENDTELNETPARMS
: A required statement block that contains Telnet parameter statements. You must create a unique TELNETPARMS block for each port and for each qualified port. The parameters define connection characteristics for the specified port. A TELNETPARMS block is required for each port that you start or modify using the VARY TCPIP,*tnproc*,OBEYFILE command.

BEGINVTAM and ENDVTAM
: A required statement block that contains Telnet mapping statements. The mapping statements define how applications and LU names are mapped (assigned) to clients. Each port that you start or modify using the VARY TCPIP,*tnproc*,OBEYFILE command must have a BEGINVTAM block. You can use one BEGINVTAM block for all ports, you can use one BEGINVTAM block for some of your ports and another BEGINVTAM block for the remaining ports, or you can use a unique BEGINVTAM block for each of your ports.

PARMSGROUP and ENDPARMSGROUP
: An optional parameter group statement in the BEGINVTAM block that contains Telnet parameter statements. The parameters define connection characteristics for the mapped clients.

See z/OS Communications Server: IP Configuration Reference for the exact syntax rules for these statement blocks. See "Using the VARY TCPIP,tnproc,OBEYFILE command to update Telnet configuration" on page 579 for information about profile processing.

Telnet initially sets all connection parameters to default values. These connection parameter values are arranged hierarchically, as shown in Figure 85 on page 575.

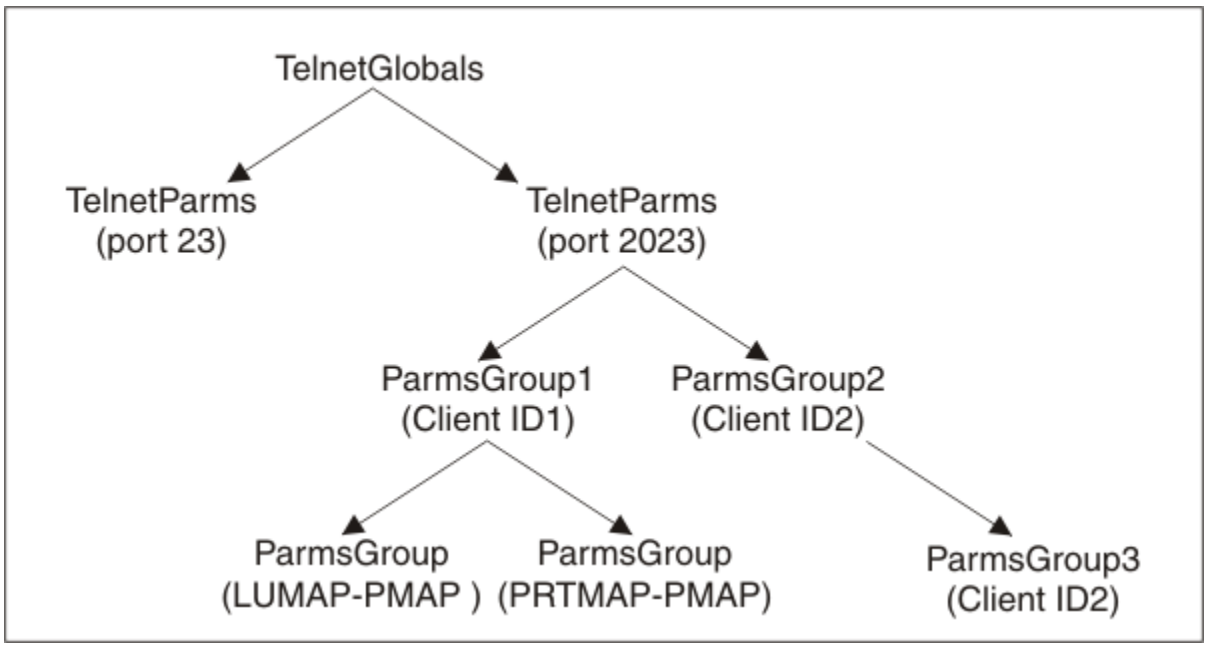


Figure 85. Telnet parameter placement

You can change these parameter values at any of the levels in the hierarchy, depending on how you want the changes to be applied to successive connection levels. Parameters that you code in the TELNETGLOBALS block are applied to all connections on all ports, unless these values are overridden by parameters in either the TELNETPARMS block or the PARMSGROUP block. Parameters that you code in the TELNETPARMS block are applied to all connections on the specified port, unless these values are overridden by parameters in the PARMSGROUP block. Parameters that you code in the PARMSGROUP block are applied to connections with clients that are mapped to that PARMSGROUP block. For a complete list of Telnet parameters, see z/OS Communications Server: IP Configuration Reference.

Eleven mapping statements are available in the BEGINVTAM block. These mapping statements map objects to client identifiers that are specified on the port. Five statements are related to application setup, four are related to LU name assignment, one maps connection parameters, and one maps monitoring rules. After a connection request is accepted, Telnet uses the mapping statements to map, or assign, as many of the eleven objects to the client as possible. This set of objects is used for the duration of the connection. For more information, see "Mapping Objects to Client Identifiers" on page 607.

The sample profile in SEZAINST(TNPROF) contains additional statements that are included as comments. These statements provide examples of advanced functions. Many of these statements are installation-specific; you need to modify these statements to suit your installation.

Steps for customizing the TN3270E Telnet server configuration data set

You can customize the TN3270E Telnet server by defining server, connection, and other configuration characteristics.

Before you begin

You need to create the data set in which you will put the configuration information. You must also specify the profile data set name on the PROFILE DD statement in the procedure JCL that is used to start Telnet. The data set must be defined with format fixed block, and must have a record length of 56-256. The block size must be evenly divisible by the record length.

If you are using multiple TN3270E Telnet servers, ensure that each server uses unique LU names, or define shared LU name groups. If you do not use unique LU names or create shared LU name groups, then the second server that uses the same LU name will not be able to establish a session; either the OPEN ACB request will fail or the cross-domain session request will fail.

For information about updating Telnet configuration, see “Using the VARY TCPIP,tnproc,OBEYFILE command to update Telnet configuration” on page 579.

Procedure

Perform the following steps to customize the TN3270E Telnet server configuration data set. A sample configuration is in SEZAINST(EZBTNPRF).

1. Define Telnet server characteristics.

   The minimum required definition is the TELNETPARMS block with the PORT statement. The following functions are optional:

   - To associate Telnet with one TCP/IP stack, see “Associating Telnet with one TCP/IP stack” on page 585.
   - To set up Telnet in a sysplex to share LU names, see “Shared LU name groups for Telnet servers” on page 586.
   - To qualify a port with the destination IP address or link name to differentiate between Telnet services, see “Qualified ports” on page 589.
   - To use multiple Telnet ports, see “Multiple ports” on page 591.
   - To reduce ECSA storage use when supporting a large number of connections, see “Reducing demand for ECSA storage” on page 654.
2. Define connection characteristics at the server or port level.
   - To determine the connection type, see “Connection mode choices” on page 592.
   - For security, see “Connection security” on page 597.
   - For persistence, see “Connection persistence” on page 604.
3. Understand the concept of mapping Objects to Client Identifiers.

   To understand the concept of mapping Objects to Client Identifiers, see “Mapping Objects to Client Identifiers” on page 607.
4. After you identify all the clients in your network, determine which clients have unique characteristics (for example, a particular client should use a certain LU name) and need to remain exact Client Identifiers, and which clients can be combined into Client Identifier groups using wildcard values.
5. Use LU name mapping statements to assign LU names to connections based on the Client Identifier. This step is required. See “LU name mapping statements” on page 617.
6. Use application mapping statements to facilitate session setup based on the Client Identifier. See “Application mapping statements” on page 621.
7. Use the connection parameters mapping statement to change connection parameters based on specific Client Identifiers. To extend connection mode, security, and persistence choices to Client Identifiers, see “Connection parameters mapping statement” on page 624.
8. Consider advanced topic features for additional Telnet functions.
   - “Advanced LU name mapping topics” on page 624
   - “Advanced application topics” on page 634
   - “Device types and logmode considerations” on page 643
   - “Using the Telnet solicitor or USS logon screen” on page 644
   - “SMF” on page 648
   - “Connection monitoring mapping statement” on page 649

Results

You know you are done when you do not receive errors after you apply the profile.

Telnet CTRACE

Telnet uses the TCP/IP stack component name SYSTCPIP. Telnet CTRACE with only the Telnet option specified provides complete information about Telnet processes. You do not need other CTRACE options for debugging most Telnet problems.

A sample Telnet component trace is supplied. The member name is CTIEZBTN in SYS1.PARMLIB. You set up tracing for Telnet in the same way that you set up tracing for the TCP/IP stack; however, there are fewer trace options needed for Telnet. You can change the component trace in the JCL by specifying a new parmlib member in the form CTIEZB*xx*. For a complete description of the trace options and for information about how to set up tracing, see the TCP/IP services traces and IPCS support information in z/OS Communications Server: IP Diagnosis Guide.

The following subset of trace options is available in the SYSTCPIP component when it is set up for Telnet:

- ALL (excludes SERIAL, STORAGE, and TIMER)
- INIT (includes OPCMDS and OPMSGS)
- MESSAGE
- MIN or MINIMUM (includes INIT, OPCMDS, and OPMSGS)
- MISC
- NONE (or OFF)
- OPCMDS (includes INIT and OPMSGS)
- OPMSGS (includes INIT and OPCMDS)
- SERIAL
- STORAGE
- TELNET
- TIMER
- WORKUNIT

Managing Telnet

This topic describes information needed to manage Telnet, such as command and port information.

Telnet commands

Telnet commands are TCP/IP commands. You must specify the procedure name on all commands; otherwise, the command is processed by the default TCP/IP stack instead of by the Telnet procedure. For example, assuming that the Telnet procedure name is TN3270E, the profile display command is as follows:

```
D TCPIP,TN3270E,PROFILE
```

If the procedure does not exist or if you incorrectly type the TCPIP keyword, then the command is assumed to be a TCP/IP command and a TCP/IP error message is displayed.

You use the following VARY and DISPLAY commands to change and monitor Telnet functions and to debug problems. Telnet VARY and DISPLAY commands are described in z/OS Communications Server: IP System Administrator's Commands.

- Use Telnet VARY commands to change the state of Telnet ports, enable or disable the use of certain Telnet LU names, and manage diagnostic tools.
  - VARY TCPIP,*tnproc*,QUIESCE,PORT blocks any new connection requests but allows existing connections to continue activity.
  - VARY TCPIP,*tnproc*,RESUME,PORT ends the quiesce state and allows new connection requests.
  - VARY TCPIP,*tnproc*,STOP,PORT ends connections on the port and closes the port.

  - VARY TCPIP,*tnproc*,OBEYFILE starts, restarts, or changes a port by updating the Telnet profile. Use the VARY TCPIP,*tnproc*,STOP and VARY TCPIP,*tnproc*,OBEYFILE commands to stop a Telnet port and then restart that port or start a new port without stopping the TCP/IP stack. You can also use these commands to increase the level of participation in the Telnet XCF group. A Telnet server that has joined the group becomes a standby LUNS, even if you used an OBEYFILE command to specify it to be a primary LUNS.

    Tip: When you issue a VARY TCPIP,*tnproc*,OBEYFILE command, the TELNETPARMS and BEGINVTAM blocks are required for each port that you start or modify.
  - VARY TCPIP,*tnproc*,ACT,*luname* activates LUs for use by the Telnet server. Specify ALL for *luname* to activate all inactive LUs with one command. This command has no effect on the VTAM state of the LU.
  - VARY TCPIP,*tnproc*,INACT,*luname* deactivates LUs for use by the Telnet server. If an LU is already in use, the command fails. This command has no effect on the VTAM state of the LU.
  - VARY TCPIP,*tnproc*,DEBUG,OFF turns off all debug activity that might have been turned on to debug a problem.
  - VARY TCPIP,*tnproc*,ABENDTRAP sets an abend trap based on unique Telnet return codes set in Telnet modules.
- Use Telnet VARY commands to change the state of a LUNS, and to enable or disable the use of certain LU names by a LUNS.
  - VARY TCPIP,*tnproc*,LUNS,START starts the takeover process by which a standby LUNS becomes the active LUNS.
  - VARY TCPIP,*tnproc*,LUNS,QUIESCE instructs a standby LUNS to stop monitoring the active LUNS. In case of a LUNS failure, this Telnet server will not be a takeover candidate. The Telnet server must be a takeover candidate for you to make changes to the LUNS using the OBEYFILE command.
  - VARY TCPIP,*tnproc*,LUNS,RESUME ends the quiesce state and instructs a LUNS to resume monitoring the active LUNS. In case of a LUNS failure, this Telnet will be a takeover candidate.
  - VARY TCPIP,*tnproc*,LUNS,ACT,*luname* activates LUs for use by the LUNS. Specify ALL for *luname* to activate all inactive LUs with one command. This command has no effect on the VTAM state of the LU.
  - VARY TCPIP,*tnproc*,LUNS,INACT,*luname* deactivates LUs for use by the LUNS. If an LU is already in use, the command fails. This command has no effect on the VTAM state of the LU.
- Use Telnet DISPLAY commands to review classic Telnet information.
  - D TCPIP,*tnproc*,PROFile displays summary or detail information about parameter statements from the TELNETGLOBALS or TELNETPARMS blocks.
  - D TCPIP,*tnproc*,OBJect displays summary or detail information about mapping statements from the Object perspective.
  - D TCPIP,*tnproc*,CLientID displays summary or detail information about mapping statements from the Client Identifier perspective.
  - D TCPIP,*tnproc*,CONN displays summary or detail information about client connections.
  - D TCPIP,*tnproc*,INACTLUS displays all LUs that were deactivated by the operator or by Telnet as a result of OPEN ACB or multilevel security problems.
  - D TCPIP,*tnproc*,STOR displays the maintenance level of a module or the amount of storage used by Telnet.
- Use Telnet DISPLAY commands to review information about Telnet in an XCF group.
  - D TCPIP,*tnproc*,XCF<,GRoup> displays the status of all members of the Telnet XCF group.
  - D TCPIP,*tnproc*,XCF,STats displays the performance status of the XCF Telnet server. If the Telnet server is a LUNS, the performance statistics between it and all LUNRs are reported.
- Use Telnet DISPLAY commands to review information about Telnet performing as a LUNS.
  - D TCPIP,*tnproc*,LUNS,OBJect displays summary or detail information about shared LUNR Objects on the LUNS.
  - D TCPIP,*tnproc*,LUNS,INACTLUS displays all the LUs that are inactive on the LUNS.

Using the VARY TCPIP,*tnproc*,OBEYFILE command to update Telnet configuration

When you use the VARY TCPIP,*tnproc*,OBEYFILE command to update Telnet configuration, new profile statements create a new configuration that is used by all connections that are accepted after you updated the file. Existing connections continue to use the configuration that was in effect when those connections were accepted. The TELNETPARMS and BEGINVTAM blocks are required for each port that you start or modify. The new configuration that you create is not a cumulative update from the previous profile. If you need to make only one change in the new profile, change the old profile or copy the profile to another data set member and make the change.

After you have updated the Telnet configuration file and the VARY TCPIP,*tnproc*,OBEYFILE command processing is completed, the new profile is labeled as the current profile, and the replaced profile becomes profile 0001. If you make another update, the new update becomes the current profile and the replaced profile becomes profile 0002. If you updated the profile for a subset of the active ports, the ports that you did not update remain unchanged. You can suppress profile debug messages by coding DEBUG OFF or DEBUG SUMMARY in the TELNETGLOBALS block and then placing the TELNETGLOBALS block in front of all other Telnet statement blocks.

New connections are associated with the current profile and use the mappings and parameters that are defined by that profile for the life of the connection. Even if you issue a VARY TCPIP,*tnproc*,OBEYFILE command to update the port, existing connections remain associated with the same profile. The statements of profiles that are not the current profile remain in effect and continue to support all connections that were established when that profile was the current profile. When all connections that are associated with a non-current profile have ended, the storage for the non-current profile mapping rules is freed and the profile is considered inactive.

The structural layout of the profiles and how connections are associated with profiles are shown in Figure 86 on page 580.

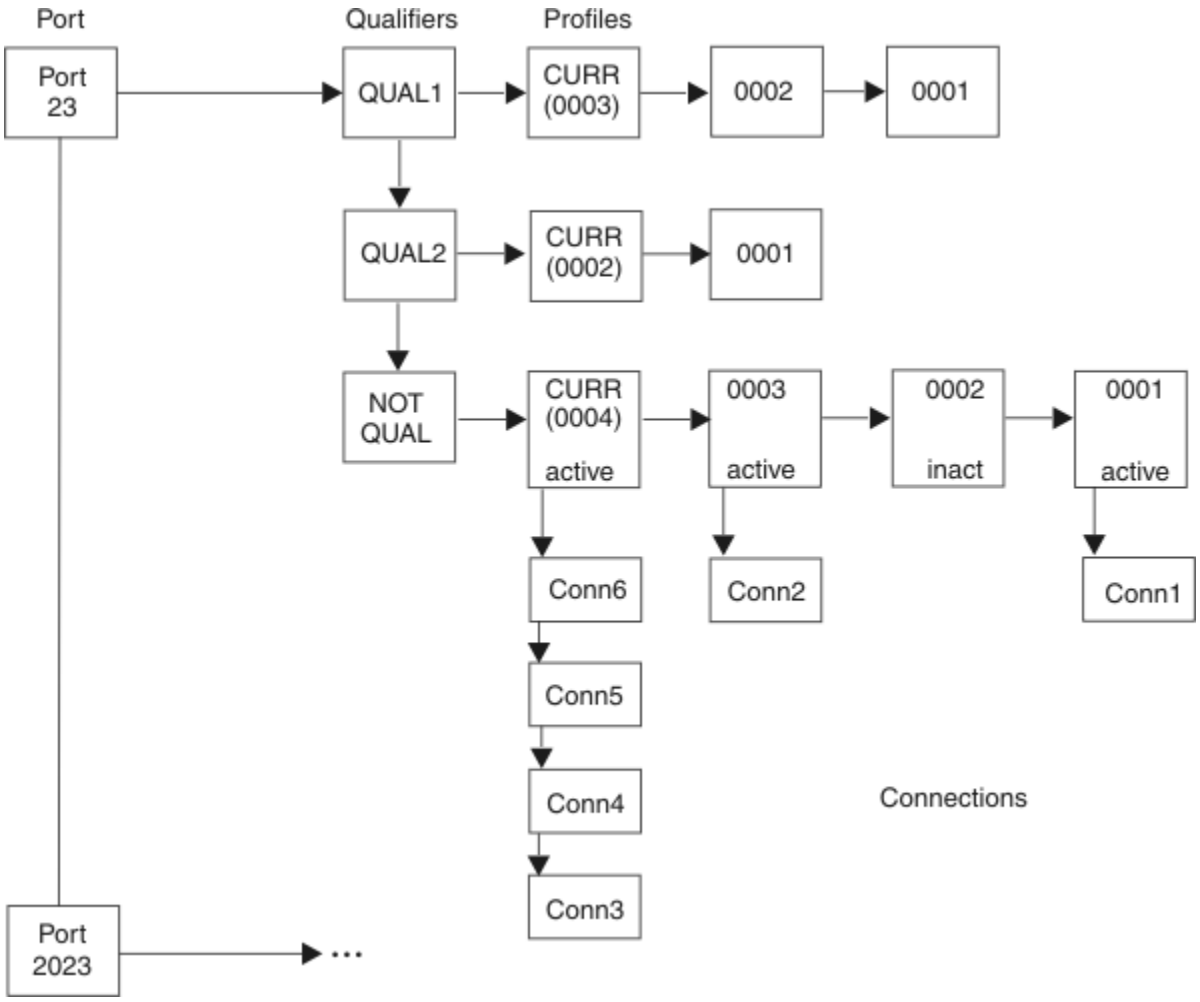


Figure 86. Telnet profiles and connections

OMVS shutdown

The TN3270E Telnet server application uses UNIX System Services; when OMVS is shut down, Telnet services are no longer available. To prevent Telnet from abnormally ending, Telnet is automatically stopped prior to OMVS shutdown, and can be restarted after OMVS is restarted.

Telnet automatically registers with OMVS; because of this registration, OMVS notifies Telnet when OMVS shutdown is requested, and OMVS waits for Telnet to stop. When Telnet is notified of the OMVS shutdown, it immediately begins to stop, as if the MVS STOP command had been issued. After Telnet is stopped, OMVS shutdown is no longer blocked by the Telnet server.

If you plan to shut down OMVS, be sure your Telnet users are logged off before you shut down OMVS.

Telnet diagnostic tools

In addition to general diagnostic tools such as CTRACE and dumps which are described in z/OS Communications Server: IP Diagnosis Guide, there are several Telnet-specific diagnostic tools available.

DEBUG messages

Telnet-specific debug messages can be turned on or off to diagnose Telnet problems related to client connections, Telnet tasks, or configuration processing. Several types of debug messages are available.

- Exception messages indicate that a problem was detected and are issued for connection, task, and configuration processing. Exception level debug is the default level.
- Summary messages indicate important status changes and are available for connection processing.

- Detail messages are issued when an important event occurs other than an error and are available for connection and task processing.
- Trace messages show detail data that is coming into and out of Telnet and are available for connection and configuration processing.

Message generation is controlled with the DEBUG statement. For information about the DEBUG statement, see z/OS Communications Server: IP Configuration Reference.

Connection processing debug options are as follows:

- Exception

  The DEBUG CONN EXCEPTION statement issues a message at the time of failure that displays the client IP address and port, connection ID, Telnet LU name, detecting module name, unique return code and brief explanation, and additional parameters if relevant. Some messages will be helpful to you and others will be helpful to IBM service. For message EZZ6035I return code details, see z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM).

  The DEBUG CONN EXCEPTION statement causes the CONN DROP message to be issued only for error conditions and inactivity reasons. DEBUG CONN EXCEPTION is the default. If more than one connection is dropped for the same reason within 15 seconds, a single message with LU name MULTIPLE will be issued. For example, if MSG07 is not coded in the DEBUG CONN DETAIL example, the connection will be dropped after the lookup failure. The CONN DROP message will include the return code and indicate that an error caused the connection drop. The following message will be produced whether or not DEBUG was coded because of the error condition.

```
EZZ6034I TNSERV11 CONN 00000011 LU TCPM1001 CONN DROP  ERR 3012
  IP..PORT: 1.12.13.14..456          EZBTPGLU
```

- Summary

  The DEBUG CONN SUMMARY statement provides tracking for the status of a connection. A summary message is written when:

  – A connection request is accepted by Telnet.
  – Connection negotiation is complete.
  – A session is established with the host application.
  – A session is dropped.
  – A connection is dropped.

  LU name, Connection ID, and client IP address and port are included in each message. In the example below a user connects to port 23. The connection is negotiated as a TN3270E connection and a session with TSO is established. The session is dropped because of client disconnect (CLNTDISC) and then the connection is dropped because of client disconnect.

```
EZZ6034I TNSERV11 CONN 00000011 LU **N/A**  ACCEPTED      23
  IP..PORT: 1.12.13.14..456
EZZ6034I TNSERV11 CONN 00000011 LU TCPM1001 NEGOTIATED TN3270E
  IP..PORT: 1.12.13.14..456
EZZ6034I TNSERV11 CONN 00000011 LU TCPM1001 IN SESSION TSO0001
  IP..PORT: 1.12.13.14..456
EZZ6034I TNSERV11 CONN 00000011 LU TCPM1001 SESS DROP  CLNTDISC
  IP..PORT: 1.12.13.14..456
EZZ6034I TNSERV11 CONN 00000011 LU TCPM1001 CONN DROP  CLNTDISC
  IP..PORT: 1.12.13.14..456
```

  In addition to tracking major state changes and providing key information, the statements can be used for diagnostic purposes. For example, a user might be attempting a connection and something is not working. The ACCEPTED, NEGOTIATED, and IN SESSION messages are major connection milestones. Using the information provided and knowing whether or not these messages are displayed can provide many clues about the connection request. The SESS DROP and CONN DROP messages give a variety of reasons about why the drop occurred.

- Detail

You can use the DEBUG CONN DETAIL statement if the DEBUG CONN SUMMARY messages do not provide enough information to solve a problem. In addition to the summary messages, the DEBUG CONN DETAIL statement issues a message at the time of failure that displays the client IP address and port, connection ID, Telnet LU name, detecting module name, unique return code and brief explanation, and additional parameters if relevant. Some messages will be helpful to the system administrator and others will be helpful to IBM service. For message EZZ6035I return code details, see z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM).

In the example below the user specified an application not in the Telnet profile and then disconnected at the client emulator.

```
EZZ6035I TNSERV11 DEBUG DETAIL CONN DETAIL
  IP..PORT: 1.12.13.14..456
  CONN: 00000011  LU: TCPM1001  MOD: EZBTPGLU
  RCODE: 3012-00  Application name is invalid.
  PARM1: 00000000 PARM2: 00000000 PARM3: 00000000

EZZ6035I TNSERV11 DEBUG DETAIL CONN DETAIL
  IP..PORT: 1.12.13.14..456
  CONN: 00000011  LU: TCPM1001  MOD: EZBTTRCV
  RCODE: 1001-02  Client disconnected from the connection.
  PARM1: FFFFFFFF PARM2: 00000461 PARM3: 00000000
```

- Trace

  The DEBUG CONN TRACE statement generates messages that contain data that was sent to and from the client, sent to and from VTAM, and sent to and from an LU name exit for a single connection. The TRACE option allows you to quickly see why a client is not connecting or why a session hangs. Be careful where you specify DEBUG CONN TRACE. Every connection that maps to the statement generates many messages. DEBUG CONN TRACE should be specified in a PARMSGROUP block that is mapped to only a few connections. Even when DEBUG CONN TRACE is specified in a PARMSGROUP block, it could flood the message console quickly if the connection is very active.

```
EZZ6034I TNSERV11 CONN 00000080 LU **N/A** ACCEPTED
  IP..PORT: 9.14.6.42..36484
EZZ6035I TNSERV11 DEBUG CONN TRACE
  IP..PORT: 9.14.6.42..36484
  CONN: 00000080  LU:           MOD: TO  CLNT
  <-C- FFFD28
  PARM1: 00000003 PARM2: 00000000 PARM3: 00000000
EZZ6035I TNSERV11 DEBUG CONN TRACE
  IP..PORT: 9.14.6.42..36484
  CONN: 00000080  LU:           MOD: FRM CLNT
  -C-> FFFB28
  PARM1: 00000003 PARM2: 00000000 PARM3: 00000000
EZZ6035I TNSERV11 DEBUG CONN TRACE
  IP..PORT: 9.14.6.42..36484
  CONN: 00000080  LU:           MOD: TO  CLNT
  <-C- FFFA2808 02FFF0
  PARM1: 00000007 PARM2: 00000000 PARM3: 00000000
```

Task processing debug options are as follows:

- Exception

  The DEBUG TASK EXCEPTION statement issues a message at the time of a failure that displays the task, detecting module name, unique return code and brief explanation, and additional parameters if relevant. Some messages will be helpful to you and others will be helpful to IBM service. For message EZZ6035I return code details, see z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM). DEBUG TASK EXCEPTION is the default for task processing debug.

  The following exception message indicates that Telnet failed to join the XCF group because the group already contains the maximum number of members allowed.

```
EZZ6035I TLUNR1   DEBUG TASK   EXCEPTION
TASK: XCF                     MOD: EZBTXXCF
RCODE: A006-01  Telnet failed to join the XCF group.
PARM1: 0000000C PARM2: 00000004 PARM3: MAX GROUPS OR MEMBERS
```

- Detail

The DEBUG TASK DETAIL statement issues a diagnostic message at certain event points that displays the task, detecting module name, unique return code and brief explanation, and additional parameters if relevant. The Detail option shows event milestones for different scenarios, enabling you to see some events of interest other than errors. Some messages will be helpful to the system administrator and others will be helpful to IBM service. For message EZZ6035I return code details, see z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM).

The following detail message indicates that an outstanding receive failed for job TLUNR1 on MVS023.

```
EZZ6035I TLUNR1   DEBUG TASK   DETAIL
TASK: XCFRECV               MOD: EZBTXRCV
RCODE: A014-01  LUNS/LUNR receive failed.
PARM1: 00000000 PARM2: 00000000 PARM3: MVS023   TLUNR1
```

Configuration processing debug options are as follows:

- Exception

  The DEBUG CONFIG EXCEPTION statement issues a message at the time of a failure that displays the line number, detecting module name, unique return code and brief explanation, and additional parameters if relevant. Some messages will be helpful to you and others will be helpful to IBM service. For message EZZ6035I return code details, see z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM). DEBUG CONFIG EXCEPTION is the default for configuration processing debug.

  The following message indicates that the value for scaninterval on line 35 is outside of the acceptable range.

```
EZZ6035I TLUNS1   DEBUG CONFIG EXCEPTION
LINE:       35              MOD: EZBTMPRP
RCODE: 805B-00  Value outside acceptable range for statement.
PARM1:          PARM2: SCANINTE PARM3:
```

- Trace

  The DEBUG CONFIG TRACE statement generates a statement message listing all words associated with the statement and a formatted control block message for each Telnet statement. This option can flood the message console if you have a very large configuration file. The CONFIG messages can be limited to a smaller, specified set of statements. If you specify a subset of statements, DEBUG CONFIG messages are issued only from those statements. If a large amount of CONFIG data is required, you can suppress the DEBUG messages by specifying the CTRACE option on the DEBUG statement. With the CTRACE option, the messages will be in the CTRACE but will not flood the console or joblog. CTRACE is the default option.

  The sample below shows data sent to the console for profile statement LUGROUP:

```
DEBUG CONFIG TRACE,LUGROUP CONSOLE

EZZ6035I TNSERV11 DEBUG CONFIG TRACE
  LINE:       82              MOD: EZBTMPRF
    LUGRP1 TCPM1094 TCPM1102
  PARM1: 00000003 PARM2: 00000052 PARM3: LUGROUP

EZZ6035I TNSERV11 TNSERV11 DEBUG CONFIG TRACE
  LINE:       82              MOD: EZBTMPRF
         E3C5D3C3 00000000 00000058 00005502 |TELC............|
         00000036 00000000 00000052 00000000 |................|
         00000000 00000000 00000000 00000000 |................|
         D3E4C7D9 D7F14040 00000000 00000000 |LUGRP1  ........|
         00000002 00000000 E3C3D7D4 F1F0F9F4 |........TCPM1094|
         E3C3D7D4 F1F1F0F2                   |TCPM1102        |
  PARM1: 00000058 PARM2: 00000036 PARM3: LUGROUP
```

The DEBUG OFF statement ensures that all debug messages are suppressed , including the exception CONN DROP messages.

The DEBUG CONN parameter can cause flooding of the operator's console. Console flooding concerns can be dealt with in several ways.

- Most DEBUG messages are, by default, assigned to routing code 11, the JOBLOG. The DEBUG option JOBLOG can be used for the same effect. However, the master console also receives routing code 11 messages by default. To stop the messages from going to the master console, issue VARY CN(01),DROUT=(11), which drops routing code 11 from the console. Another DEBUG option, CONSOLE, will direct the messages to the master console, routing code 2, and the teleprocessing console, routing code 8. The CTRACE option suppresses messages completely while continuing tracing at the debug level.
- If DEBUG messages are being used primarily for problem diagnosis, the VARY TCPIP,*tnproc*,OBEYFILE command can be used to keep the number of messages low. Start Telnet initially without DEBUG coded. When a problem appears, issue a VARY TCPIP,*tnproc*,OBEYFILE command for a Telnet profile that includes the DEBUG statement. Only new connections to the new profile will produce messages. After data is obtained, issue another VARY TCPIP,*tnproc*,OBEYFILE command for a Telnet profile that omits the DEBUG statement.
- If the Client Identifier of the client having the problem is known, include DEBUG in a PARMSGROUP statement. Using PARMSMAP, map that group to the client. Debug messages for only that client will be issued.

The VARY TCPIP,*tnproc*,TELNET,DEBUG,OFF command can be issued to turn off DEBUG for all connections associated with all profiles, including the current profile. It will also turn off TASK and CONFIG DEBUG messages. To turn on DEBUG again, issue a VARY TCPIP,*tnproc*,OBEYFILE command with the Debug option specified in the Telnet profile. Summary messages for CONN DROP due to errors or time-outs will also be suppressed. Use DEBUG EXCEPTION to retain these messages.

MSG07

The MSG07 parameter is very helpful when diagnosing problems. It allows Telnet to send a message to the client indicating an error occurred and what the error was. Something simple like a mistyped application name can be corrected by the user without additional help. Even for more difficult problems, MSG07 provides a good starting point. It is recommended that MSG07 always be coded unless there are reasons not to send error messages to the client.

Abend trap

The VARY TCPIP,*tnproc*,TELNET,ABENDTRAP,*module*,*rcode*,*instance* command can be used to set up an abend based on the variables specified. Abend trap has three variables:

module is required. It is the module detecting the error. It can have a wildcard value by using asterisk (*) at the end. If a single * is used, any module reporting the specified return code will cause an abend. The module name "OFF" turns off an active trap.

rcode is optional. It is the return code reported and cannot have a wildcard value.

instance is optional. It is the instance of the return code and cannot have a wildcard value.

Below is an example setting the abend trap and then issuing a profile display to verify the trap is set. In the example, when EZBTTRCV reports an error code of 1001, Telnet will issue an abend with reason code '3133'x. The state of the trap changes from "ACTIVE" to "TRIPPED". No more abends will be issued. After being tripped, the abendtrap command must be issued again to activate the trap. While the trap is active, the abend trap can be turned off by specifying "OFF" as the module name. An active trap cannot be changed directly. The current trap must be tripped or turned off before a new command is accepted.

```
V TCPIP,TCPCS6,T,ABENDTRAP,EZBTTRCV,1001
EZZ6038I TCPCS6 COMMAND ABENDTRAP EZBTTRCV COMPLETE

D TCPIP,TCPCS6,T,PROF
EZZ6060I TCPCS6 PROFILE DISPLAY
  PERSIS   FUNCTION    DIA  SECURITY  TIMERS  MISC
 (LMTGQAK)(OATSKQSWHT)(DRF)(PCKLECXN)(IKPSTS)(SMLT)
  -------  ----------  ---  --------  ------  ----
  LM*R*P*  **TSBQ*WHT  TJ*  SSH*ESX*  ***STS  SDD*
----- PORT:    23  ACTIVE           PROF: CURR CONNS:      1
-------------------------------------------------------------
```

```
   ABENDTRAP       : EZBTTRCV  1001  FF  ACTIVE
4 OF 4 RECORDS DISPLAYED
```

TESTMODE

You can use the TESTMODE parameter statement in a TELNETPARMS block to allow a port to be processed by Telnet but then dropped before the statements for that port become active. Using TESTMODE ensures that LU assignments, security levels, and other Telnet parameters are not compromised due to syntax errors.

Displays

Use the DISPLAY commands to verify that the configuration is what you thought it was, that connections are mapping to the correct parameters and Objects, and to check XCF Telnet LUNS and LUNR status and performance. For details about the commands available, see “Managing Telnet” on page 577 and z/OS Communications Server: IP System Administrator's Commands.

Tracing

If you cannot resolve a problem using the available tools, IBM service will likely need a CTRACE with option Telnet. For details, see “Telnet CTRACE” on page 577. You might also need to activate one of the following additional traces:

- Full data trace

  If the problem is data related, use the FULLDATATRACE statement to trace all the data coming into and leaving Telnet rather than tracing only the first 64 bytes of data. FULLDATATRACE will cause a trace-wrap condition more quickly so it should be set only if needed. It should be set in PARMSGROUP instead of TELNETPARMS if a subset of clients can be identified.

- Telnet subagent trace

  For Telnet SNMP subagent problems, use the TNSATRACE keyword on the TNSACONFIG statement in PROFILE.TCPIP at startup. This will generate trace points throughout Telnet subagent processing, in addition to tracing data passed between the Telnet subagent, Telnet, and the TCP/IP stack. Subagent tracing can also be enabled after Telnet has been started by using the VARY TCPIP,*tnproc*,OBEYFILE command. To enable tracing using the VARY TCPIP,*tnproc*,OBEYFILE command, the subagent must first be disabled and then re-enabled with the TNSATRACE keyword. Trace data is written to the syslog daemon. Subagent tracing can be disabled using the NOTNSATRACE keyword.

- DBCS trace

  If the problem is with a DBCS connection, use the DBCSTRACE statement in the TELNETPARMS block or in the PARMSGROUP block to produce DBCS-specific trace entries in the SYSPRINT and TNDBCSER data sets.

Telnet configuration data set customization details

This topic includes details for “Steps for customizing the TN3270E Telnet server configuration data set” on page 575.

Associating Telnet with one TCP/IP stack

In an INET environment where only a single TCP/IP stack can be running, Telnet client connections are automatically associated with the active stack, and you do not need to explicitly associate the Telnet server to the stack.

In a common INET (CINET) environment, Telnet is associated with all stacks that are running. When multiple stacks are supporting a single Telnet server, Netstat displays might not display all Telnet connections. The only connections that are displayed are connections that are supported by the stack from which the command is issued. If another stack is started while Telnet is active, the current LISTEN

for the port is cancelled and is reissued automatically to include the new stack. If you explicitly associate Telnet to one TCP/IP stack, all Telnet clients must connect through that stack.

If you want to explicitly associate Telnet with one stack for control purposes or for functionality support, use the TCPIPJOBNAME statement in the TELNETGLOBALS block when you start Telnet. If you use the TCPIPJOBNAME statement, you must continue to use it on all future profile updates to set affinity to the same stack.

The Telnet SNMP subagent requires that you explicitly associate Telnet with a single TCP/IP stack. Telnet SNMP subagent activation requires that you register the stack name with the agent. If you do not specify the TCPIPJOBNAME statement, then Telnet blocks the subagent activation request. The Telnet SNMP subagent can register with only one agent, and each agent can support only one Telnet subagent. If you are going to use the Telnet SNMP subagent, plan for one agent per Telnet subagent. If multiple Telnet SNMP subagents initialize to the same agent, the agent forwards all data requests to the first subagent that connected, and all other initialization attempts are queued. If the first subagent ends, then the next subagent in the queue receives all data requests.

Shared LU name groups for Telnet servers

SNA architecture requires every LU in a VTAM network to have a unique name. Multiple Telnet servers create added administrative effort to ensure that LU names are unique among the servers. However, if your environment uses multiple Telnet servers running on a single system or in a sysplex, then you can designate one Telnet server to be the LU name server (LUNS). The LUNS manages LU name assignments from LU groups among the group of Telnet servers, each known as an LU name requester (LUNR).

Shared LU groups are defined at each LUNR and sent to the LUNS. Shared LU group definitions can be the same or different at each LUNR. The LUNS allocates an LU name to a particular LUNR only if that LUNR defined the LU in a shared LU group. The LUNS manages LU names by ensuring that only one LUNR at a time is using a particular LU name. You can use load balancing to distribute Telnet client connections across several LUNR Telnet servers that have identical shared LU name configurations.

A single Telnet server can support both shared and unshared LU groups. Existing unshared LU group definitions continue to be managed at the local Telnet level.

You can configure a Telnet server to be only a LUNS, or a Telnet server can be a LUNS and also function as a regular Telnet server. Running Telnet as only a LUNS in its own address space has the following advantages:

- Telnet port server functions will not compete with the LUNS for resources within the address space.
- You can separate Telnet roles, which makes problem diagnosis easier.
- You can stop and restart the Telnet LUNS without stopping the Telnet ports, and you can stop and restart the Telnet port servers without stopping the Telnet LUNS.
- You can set the Telnet LUNS priority to a different priority than that of Telnet port servers.

Telnet uses XCF local group services to define the set of Telnet servers that participate in a shared LU name management group. You specify the level of participation of a particular Telnet server with shared LU name management by specifying or omitting the XCFGROUP block. For configuration details, see “Steps for defining a LUNS and a LUNR” on page 588.

- Classic Telnet server

  If you do not specify the XCFGROUP block for a server or you specify NOJOIN in the XCFGROUP block, the server does not join the XCFGROUP and cannot participate in shared LU name management.

  – The server does not join an XCF group
  – The server is not included in the Telnet XCF,GROUP display
  – Shared LU group Objects cannot be defined on the server

  This is the default behavior of a Telnet server.
- XCF Telnet server

If you do specify the XCFGROUP block for a server and you specify JOIN in the XCFGROUP block, the server joins the Telnet XCF group and is able to participate in shared LU name management.

  - The server is included in the Telnet XCF,GROUP display that can be issued for any XCF Telnet server that is in the same sysplex group
  - The health of the Telnet server is monitored by XCF
  - The server supports shared LU group Objects
  - The server can be an LU name requester (LUNR)

    The LUNR is in standby state; it becomes active when shared LU group Objects are found while processing a profile. If you issue a VARY OBEYFILE command and there is no active profile that has shared LU group Objects, then the LUNR reverts to standby state.

- XCF LUNS Telnet server

  If you specify the XCFGROUP block for a server and include the appropriate statements, you can configure the server to have LUNS capability.

  - The server has all the characteristics of an XCF Telnet server
  - The server is an LU name server (LUNS)

After you have configured the level of participation, you can use the VARY OBEYFILE command to move the Telnet server to a higher level; however, you cannot use the VARY OBEYFILE command to move the server to a lower level.

There can be only one active LUNS in a sysplex; that LUNS manages requests for shared LU names so that the names are used by only one Telnet server at a time. The remaining Telnet servers are LU name requesters (LUNRs) that request LU names from the LUNS. The LUNR Telnet server connects to an administration listener socket that is created by the LUNS Telnet server. This TCP connection transmits all LU name requests and responses; this connection is not used for status signalling between the XCF Telnet servers. XCF services including XCF user state field updates, XCF status monitoring, and XCF group events are used to communicate LUNS and LUNR state changes and health.

You can configure a LUNS as a primary or a backup LUNS. When the Telnet server is started, a primary LUNS checks to determine whether there is already an active LUNS in the XCF group. If there is not, that primary LUNS attempts to become the active LUNS. If there is already an active LUNS, the primary LUNS becomes a standby LUNS. If several Telnet servers that are designated as a primary LUNS are started concurrently, then the race winner becomes the active LUNS and the others become standby. A backup LUNS always becomes a standby LUNS.

When an active LUNS fails, the following sequence of events automatically occurs:

1. All the LU name servers that are in standby mode examine the list of LU name servers that are in standby mode.
2. The standby LUNS that has the highest configured rank, regardless of whether that LUNS was started as a primary or backup LUNS, automatically becomes the active LUNS.

   If several standby LU name servers have the same rank, then all those servers attempt to take over. The race winner becomes the new active LUNS and the others return to standby.

You can initiate manual takeover of a Telnet LUNS at any time. You can direct an operator command to any LUNS that is in standby or joined mode and tell it to start and become the active LUNS. The current LUNS will change to standby mode.

You can define all the Telnet servers in a sysplex that participate in shared LU name management to join the same XCF group. However, if your environment has one or more of the following characteristics, then you should partition the sysplex into Telnet subplexes:

- All Telnet servers in a Telnet XCF group must have IP connectivity to each other. If your sysplex is partitioned into TCP/IP subplexes that do not have connected routes to each other, then you should partition the sysplex into Telnet subplexes along the same TCP/IP boundaries.

- All Telnet servers in a Telnet XCF group should run on VTAMs in the same network. If your sysplex is partitioned into VTAM subplexes and the VTAMs are in different networks, then you should partition the sysplex into Telnet subplexes along the same VTAM network boundaries.
- If you use several Telnet ports, each with a large set of shared LU names that do not overlap, and you load balance these ports across several Telnet servers, you might want to partition the sysplex into Telnet subplexes along the Telnet port boundaries. Subplexes reduce the shared LU name management workload on the LUNS and provide operational independence of the LUNS. All Telnet servers that are DVIPA targets of the same distributed port should be members of the same Telnet subplex.

A static VIPA is the best type of IP address to use for the LUNS administrative listener socket. Dynamic VIPA can work for the LUNS administrative listener, but there are cases in which dynamic VIPA can cause confusion.

- If VIPADEFINE moves the IP address away from the LUNS, LUNRs already connected to the LUNS will continue to be able to send requests and receive replies. However, a new LUNR attempting to connect will be directed to the new TCP/IP stack, and the LUNS will not be there.
- VIPARANGE will work if you can guarantee that each LUNS is supported by a different TCP/IP stack, and that the LUNS is the application that creates the IP address. If not, and the creating application ends, the LUNS socket will be closed. A standby LUNS on the same stack would be a problem. The standby LUNS sets up its listener before the original LUNS finishes cleanup. When the original LUNS finishes, the new LUNS listener socket will be closed.

For information about the XCFGROUP block, see z/OS Communications Server: IP Configuration Reference. For information about Telnet LUNS display, XCF display, and LUNS VARY commands, see z/OS Communications Server: IP System Administrator's Commands.

Steps for defining a LUNS and a LUNR

If your environment uses multiple Telnet servers running on a single system or in a sysplex, you can designate one Telnet server to be the LU name server (LUNS). The LUNS ensures that LU names are unique among your Telnet servers, managing LU name assignments.

Before you begin

Determine which LU names are going to be shared and determine the setup of the LU groups. If you are running the Telnet server on multiple systems, ensure that all systems can function in a sysplex. The systems must have at least an XCF group connection. You do not need a coupling facility.

Procedure

Perform the following steps to configure an LU name server (LUNS) or an LU name requester (LUNR):

1. On the XCFGROUP statement block on the TELNETGLOBALS statement block, define the level of participation in the Telnet XCF group and in shared LU name management.
   - Specify the NOJOIN parameter (or do not include the XCFGROUP block) if this Telnet server should not join the Telnet XCF group. If you specify NOJOIN (or omit the XCFGROUP block), the server is not included in the XCF,GROUP display and does not allow the definition of shared LU group objects.
   - Specify the JOIN parameter if this Telnet server should join the Telnet XCF group. JOIN is the default if you code XCFGROUP. If you specify JOIN or it is in effect by default, the server is included in the XCF,GROUP display that is issued by any other XCF Telnet server; the server has LUNR capability and supports shared LU group objects. With additional XCFGROUP statements, the server can have LUNS capability.
2. Determine whether you need to partition your Telnet server sysplex into Telnet subplexes.
   - If you are not partitioning your sysplex into subplexes, do not code the SUBPLEX parameter on the XCFGROUP statement. All Telnet servers in the sysplex that participate in shared LU name management join the default Telnet XCF group EZZTLUNS.
   - If you are partitioning your sysplex, configure the SUBPLEX parameter with a suffix that is 1 - 4 characters in length. The specified suffix is right-aligned and overlays the end of the string EZZTLUNS

to form a unique subplex string. For example, if the suffix value is 23, Telnet joins XCF group EZZTLU23.

3. Determine the frequency with which you want to monitor the health of the Telnet LUNS and LUNR.

   Use the XCFMONITOR parameter to set the frequency. At the specified time interval, Telnet checks the health of the LUNS, LUNR, and XCF Telnet tasks, and checks the health of the connection between the LUNS and LUNR. If any of these tasks or the connection appear to be unresponsive, message EZZ6099I is issued and the X indicator is set under the PDMON column in the XCFGROUP display. No action is taken by Telnet to correct the issue. A short time value can lead to false indications in a busy system. A long time value might not provide feedback quickly enough so that you can take appropriate action. The default value is a good compromise between these two possibilities.

4. To configure a LUNR, perform the following steps:

   a) Configure the following parameters on the XCFGROUP statement:

      - CONNECTTIMEOUT: You can use this parameter for a LUNR only. Specify the length of time that a LUNR attempts to establish a connection to the LUNS before quiescing its LUNR capabilities. If the CONNECTTIMEOUT time elapses and the LUNR has not been able to connect with the LUNS, then the LUNR drops all connections that are waiting in negotiation for an LU name and quiesces all ports that have shared groups. Dropping connections that are in negotiation state enables the client to try again to connect to Telnet. Quiescing the port alerts a distributor that it should send requests to other working Telnet LUNRs. When the client connection is dropped and the client reconnects to the distributor, the connection is directed to a working LUNR. When the LUNS-LUNR connection is established, the port is automatically resumed and new client connections are again routed to this LUNR. Setting the CONNECTTIMEOUT time to 0 causes client connections to remain in LU name negotiation until LUNS-LUNR communication is established.
      - RECOVERYTIMEOUT: You can use this parameter for a LUNR only. Specify the length of time that the LUNR attempts to establish a connection when the LUNS is in RECOVER state before dropping active client connections that have shared LU names assigned to them. A new LUNS in RECOVER state cannot become active until it has received updates from all LUNRs that have shared LUs that are in use. If a LUNR cannot communicate with the LUNS during recovery time and the LUNR has shared LUs assigned, then the LUNS cannot become active and none of the LUNRs are able to receive LU name assignments. If the LUNR drops all active client connections with shared LUs, then the LUNS is alerted through XCF that the LUNR no longer owns shared LUs, and the LUNS can then become active.

   b) Define the shared LU groups for the LUNR.

      Use the SLUGROUP and SPRTGROUP statements to define the shared LUs.

5. To configure a LUNS, specify LUNS on the XCFGROUP statement and configure the following parameters on the LUNS statement:

   - *ipaddr port*: Specify the IP address and port on which the LUNS will listen for its administrative connection to communicate with LUNRs.
   - PRIMARY or BACKUP: Specify whether this Telnet server is a primary LUNS or a backup LUNS.
   - RANK *nnn*: Specify the start rank of this LUNS, relative to others, when this LUNS is in standby mode and the active LUNS fails.

Results

You have configured your Telnet servers correctly when you can issue a D TCPIP,*tnproc*,XCF,GROUP command and see the appropriate LUNS or LUNR status for each server. You should be able to connect a client to the LUNR and at the LUNS you should be able to issue a D TCPIP,*tnproc*,LUNS,OBJ command to verify that an LU was allocated from the LUNS.

Qualified ports

In some cases all clients need to use the same port number, but the Telnet parameters need to be differentiated by destination IP address or destination linkname. The destination IP address can be either an IPv4 or IPv6 IP address.

For example, two TN3270E Telnet servers are going to be merged into one server. Currently, server 1 is bound to IP address 1.1.1.1 and is running with a set of definitions for port 23. Server 2 is bound to IP address 2.2.2.2 and is running with a different set of definitions for port 23. Before the servers are merged into one, users connect to either 1.1.1.1,port 23 or 2.2.2.2,port 23, depending on which Telnet services the users want. The sample definition statements are:

```
Server 1

 TelnetParms
  Port 23
  Inactive 600      ; Drop after 10 minutes of no activity
 EndTelnetParms

 BeginVTAM
  Port 23
  DefaultLus TCPABC01..TCPABC49 EndDefaultLus
  DefaultAppl TSO
 EndVTAM

Server 2

 TelnetParms
  Port 23
  Inactive 0        ; Never drop
 EndTelnetParms

 BeginVTAM
  Port 23
  DefaultLus TCPABC50..TCPABC99 EndDefaultLus
  DefaultAppl CICS
 EndVTAM
```

After the servers are merged, both home addresses are supported by a single server. One way to keep the Telnet definitions separate is to change the port number in one of the definition sets. For instance, the port 23 definitions associated with the old server 2 could be changed to be port 2023. The end result is one TN3270E Telnet server with port 23 and port 2023, where port 23 has the definitions used in the old server 1 and port 2023 has the definitions used in the old server 2. The definitions are still separate. However, all the users who were connecting to 2.2.2.2 port 23 now have to change their clients to port 2023. The sample definition statements are changed as follows:

```
Merged Server

 TelnetParms
  Port 23
  Inactive 600      ; Drop after 10 minutes of no activity
 EndTelnetParms

 BeginVTAM
  Port 23
  DefaultLus TCPABC01..TCPABC49 EndDefaultLus
  DefaultAppl TSO
 EndVTAM

 TelnetParms
  Port 2023
  Inactive 0        ; Never drop
 EndTelnetParms

 BeginVTAM
  Port 2023
  DefaultLus TCPABC50..TCPABC99 EndDefaultLus
  DefaultAppl CICS
 EndVTAM
```

With port qualification, the system administrator can qualify the port number with the destination IP address or linkname to keep the Telnet services separate. In this case, the destination IP address is used. The qualified port allows the users of either old stack to connect without making any changes to their client. The sample definition statements would be changed to:

```
Merged Server

 TelnetParms
  Port 23,1.1.1.1
```

```
 Inactive 600      ; Drop after 10 minutes of no activity
EndTelnetParms

BeginVTAM
 Port 23,1.1.1.1
 DefaultLus TCPABC01..TCPABC49 EndDefaultLus
 DefaultAppl TSO
EndVTAM

TelnetParms
 Port 23,2.2.2.2
 Inactive 0        ; Never drop
EndTelnetParms

BeginVTAM
 Port 23,2.2.2.2
 DefaultLus TCPABC50..TCPABC99 EndDefaultLus
 DefaultAppl CICS
EndVTAM
```

You cannot QUIESCE, RESUME, or STOP a qualified portion of a port. If the port has several qualified port profiles, the VARY TCPIP,*tnproc*,QUIESCE, the VARY TCPIP,*tnproc*,RESUME, and the VARY TCPIP,*tnproc*,STOP commands affect all qualified port profiles associated with the port being quiesced, resumed, or stopped. In the example above, V TCPIP,*tnproc*,T,STOP,PORT=23 will stop port 23,1.1.1.1 and port 23,2.2.2.2. It is not possible to stop port 23,1.1.1.1 or port 23,2.2.2.2 individually. All display commands that allow port specification allow you to specify a qualified port. If just the port number is specified, only the unqualified port, if it exists, is displayed. The qualified port profiles are not displayed. DBCSTRANSFORM can be active on only one port, but can be active on one, some, or all of the qualified profiles of that port.

Multiple ports

Telnet supports up to 255 ports on one server. A unique TELNETPARMS block must be created for each port or qualified port. Telnet allows the use of the same BEGINVTAM block for all ports, some ports, or a unique BEGINVTAM block for each port. Both TELNETPARMS and BEGINVTAM blocks are required for each port started or modified by a VARY TCPIP,*tnproc*,OBEYFILE command. There are several reasons that more than one Telnet port or qualified port might be needed. The most common reasons are to simplify the setup of clients on the workstation and the logon process, and to differentiate client security needs.

Assigning a single application to a port simplifies the setup of clients on the workstation and the logon process. Workstation clients can be labeled with the associated application name and then be set up to connect to the appropriate port or qualified port. With a client per application on the workstation, the user can select the needed client, connect, and be immediately in session with the application defined on the DEFAULTAPPL statement in BEGINVTAM. This implementation requires a unique BEGINVTAM block for each port due to the unique DEFAULTAPPL statements. The example below shows how to set up TSO, IMS, and CICS on ports 23, 2023, and 4023, respectively. The same LU names are used in each BEGINVTAM block. Telnet maintains a master LU "in-use" registry across all ports so that the same LU name will not be used by two different ports.

```
TELNETPARMS
      PORT  23
ENDTELNETPARMS
TELNETPARMS
      PORT  2023
ENDTELNETPARMS
TELNETPARMS
      PORT  4023
ENDTELNETPARMS

BEGINVTAM
      PORT  23
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      DEFAULTAPPL   TSO
ENDVTAM
BEGINVTAM
      PORT  2023
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      DEFAULTAPPL   IMS
ENDVTAM
```

```
BEGINVTAM
      PORT  4023
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      DEFAULTAPPL   CICS
ENDVTAM
```

Assigning different security levels to different ports is an easy way to differentiate client security needs. External connections might require SSL security, while internal connections do not. Other than that difference, all other aspects of the Telnet profile can be the same. For example, external clients can connect to port 23 of a firewall that converts the request to the Telnet secure port 992. Internal clients would connect directly to the Telnet basic port 23. The statements below show how two ports allow implementation of different security levels. Note the same BEGINVTAM block is used for both ports, which can significantly reduce profile maintenance complexity. The PORT statement in BEGINVTAM links the BEGINVTAM block to the multiple TELNETPARMS blocks defined.

```
TELNETPARMS
      PORT  23
ENDTELNETPARMS
TELNETPARMS
      TTLSPORT  992
ENDTELNETPARMS
BEGINVTAM
      PORT  23 992
      DEFAULTLUS TCPABC01..TCPABC99 ENDDEFAULTLUS
      ALLOWAPPL   *
ENDVTAM
```

If a profile that contains a new port number is processed, it is treated as an additional port, and the VARY TCPIP,*tnproc*,OBEYFILE command request will succeed if all parameters for the new port are correctly specified. Existing, non-referenced ports remain active and unchanged. You can use the VARY TCPIP,*tnproc*,TELNET,STOP command to stop a port.

Connection mode choices

Telnet supports several connection types. The negotiation process is hierarchical in the order listed below:

- TN3270 Enhanced (TN3270E)
- TN3270
- Linemode
  - Standard
  - Binary
  - Transform

TN3270E is the default connection mode for Telnet. If the client refuses TN3270E mode, Telnet tries TN3270 mode. If the client refuses TN3270 mode, Telnet tries Linemode. Telnet does not support Network Virtual Terminal (NVT) mode, except to allow the negotiation of TN3270E, TN3270, or linemode connections.

Note: The Type of Service (ToS) byte, also known as the Differentiated Services field, is not managed directly by Telnet. If you want to use Differentiated Services for Telnet, use the Quality of Service (QoS) support discussed in Chapter 15, “Quality of service,” on page 831.

TN3270E and TN3270 are very similar. If the TN3270E functions are not needed, the user does not notice any difference between TN3270E and TN3270 connections. In some cases, older clients do not properly refuse the server request for a TN3270E connection, and the connection is dropped. In these unusual cases, use the NOTN3270E parameter to disable the TN3270E function for those clients. Similarly, use the NOSNAEXT parameter for any client that does not properly negotiate the extension functions (Contention Resolution and SNA Sense). TN3270E/NOTN3270E and SNAEXT/NOSNAEXT parameters can be coded at all three parameter block levels for different levels of granularity.

TN3270E and TN3270 clients can receive a Telnet solicitor screen to submit an application name, User ID, and password or password phrase to Telnet. The cursor is positioned on the application line unless

the OLDSOLICITOR parameter is specified which causes the cursor to be positioned on the user line. See “Using the Telnet solicitor or USS logon screen” on page 644 for detailed information.

The ATTN key function is supported over TN3270, TN3270E, and Transform Linemode connections. It is not supported over Standard or Binary Linemode. Default LOGMODEs for TN3270E connections are SNA, and default LOGMODEs for TN3270 and Transform connections are non-SNA. Telnet processes the ATTN key differently for SNA and non-SNA LOGMODEs. In addition, Telnet can be configured to handle double ATTNs sent by some clients by specifying SINGLEATTN. See “Device types and logmode considerations” on page 643 for more information.

For TN3270E, LU assignment is done during connection negotiation. For TN3270, LU assignment is done at application selection time. To delay LU assignment until application selection time for TN3270E, specify the SIMCLIENTLU parameter. See “LU mapping by application name” on page 630 and “LU mapping selection rules” on page 631 for details.

You might experience unexpected results if you start a Telnet session from within an application that is already connected using Telnet. For example, if you start a new Telnet session from within a TSO session that was established on a TN3270E connection, the keyboard will unlock when it seems it should not. This happens when an unlock keyboard intended for only the original, first session is sent from Telnet. The second session should remain locked but does not. An unlock keyboard intended for only the first session has the affect of unlocking the keyboard for both the first and second session because both are represented by the same client.

Some host applications send 3270 read commands (for example, X'F2' read buffer) to the client during the course of a session. Telnet sends an unlock keyboard sequence (that is, X'F1C2') before the read command is sent to the client. This is the default behavior or can be specified by coding UNLOCKKEYBOARD BEFOREREAD. In some cases, a problem can arise if the keyboard is unlocked prior to the read command being forwarded to the client. The unlock keyboard sequence allows transmission of buffered keyboard data to the host application. The buffered keyboard data is not expected in response to a read command. The UNLOCKKEYBOARD AFTERREAD parameter can be used to send the unlock keyboard sequence after the read command rather than before. In most cases, the default value will suffice and there is no need to code or change the setting of this parameter. Certain applications, however, will issue error messages when buffered keyboard data is unexpectedly received from the client. In these cases, UNLOCKKEYBOARD AFTERREAD can be coded to resolve the application error.

Some applications expect the user to initiate session data traffic. If a USSMSG10 screen or solicitor screen was used to initiate the session, the keyboard is locked. A BIND flows to the client on a TN3270E connection alerting the client to unlock the keyboard. A non-TN3270E connection does not support sending a BIND to the client. Therefore, when a BIND is received from the application, Telnet sends an unlock keyboard to the client on a non-TN3270E connection to ensure the user can initiate data traffic if necessary. This behavior is the default or can be specified by coding UNLOCKKEYBOARD TN3270BIND. In some cases, the unlock keyboard might not be correctly interpreted by an older client. If this is the case, UNLOCKKEYBOARD NOTN3270BIND can be coded to stop Telnet from sending an unlock keyboard when a BIND is received.

The two unlock keyboard functions must be specified on a single UNLOCKKEYBOARD statement. If only one is specified, it is assumed the other is the default value. The UNLOCKKEYBOARD parameter can be coded at all three parameter block levels for different levels of granularity.

TN3270 Enhanced

TN3270 Enhanced (TN3270E) connections support full-screen 3270 emulation that is sometimes referred to as TN3270 Extended. Do not confuse TN3270E function with the IBM 327x device types that end in -E (for example, 3278-2-E). In these cases, the *E* indicates that the terminal supports Extended field attributes such as color and highlighting and is not related to Telnet functions.

Telnet is often used as the primary method of connection between client workstations and the SNA mainframe environment. To make this form of remote connection as seamless as possible, Telnet terminal emulation simulates actual SNA terminals as closely as possible. To accomplish this, RFC1647 and RFC2355 (both known as TN3270E) add the ability to specify device names at connection time, add

support for printer devices, and add additional SNA functions. An Internet draft, RFC 2355 Extensions, adds Contention Resolution and SNA Sense code support.

Device name specification
Telnet assigns LUs based on the LU mapping statements supplied. Clients are assigned a device name (Telnet LU name) based on those statements. However, a TN3270E client can optionally specify that a particular device name be assigned, or it can specify that a device name from a pool of LUs be assigned. If the specified device name is allowed for this client based on the LU mapping statements and the LU is available, Telnet assigns the specified device name. If the specified device pool is allowed for this client based on the LU mapping statements and an LU within the pool is available, Telnet assigns a device name from the specified pool. Otherwise, the request is rejected with an appropriate reason code, and the connection is dropped. See "Mapping Objects to Client Identifiers" on page 607 for additional LU mapping information.

328x printer support
Many Telnet clients emulate 328x class printers (device type IBM-3287-1). Most support both SNA Character Stream (SCS) as an LU1 and 3270 data stream as an LU3. The support of each is negotiated at connection time. When connected in TN3270E mode, Telnet supports these emulators in a manner similar to terminal LUs. Telnet can be configured to initiate a session at connection time or to open an ACB to let the application initiate the session. The bind initiating each session is sent to the client, and the bind informs the emulator which data stream to expect. The VTAM application perceives the Telnet LU to be an an actual 3287-class printer and sends the SCS or 3270 data to the Telnet LU. Telnet passes the data on to the client, which prints the data. Telnet printer support allows you to use a single product, Telnet, to control both SNA terminals and SNA printers.

Some Telnet client printer emulators can request to be associated with a terminal device name by specifying the terminal device name during connection negotiation. Using printer association, users can connect their Telnet terminals to an application and then have Telnet assign an associated printer device name based on the terminal name. To associate printers with terminals, Telnet must have a printer device pool of LUs defined and a terminal device pool of LUs defined with each having the same number of device names. For more information, see "Associated printer function" on page 627.

Additional negotiated TN3270E support
Responses and SysReq functions are supported by most clients that support TN3270E connections. Contention Resolution and SNA Sense support are newer and less prevalent.

- Responses - The client or host VTAM application can request that it receive definite, exception, or no response. Client responses to application requests provide more accurate response information for application-based monitoring tools compared to TN3270 connections. When the client negotiates responses and the SNA application sends data with a definite response, the Telnet Server disables the DELAYACKs function at times. For TN3270 connections, Telnet must intercept response requests from the host and respond on behalf of the client, incorrectly reducing measured response time. Telnet monitoring will provide accurate response time information for either TN3270E or TN3270 connections.
- SysReq function - The user can request that a current session be dropped by entering LOGOFF (in upper, lower, or mixed case) after pressing the SysReq key. If LUSESSIONPEND is not mapped to the client, the connection will be dropped. Otherwise, a USSMSG10 screen is sent to the client. If, instead of entering LOGOFF, the SysReq key is pressed a second time and if the application supports LUSTAT 082B (presentation screen is lost), the previous screen is resent to the client emulator.
- Contention Resolution - Improves communication between the client and host VTAM application about which owns the send state. Contention Resolution includes the following items:
  - Keyboard Restore Indicator (KRI) - Whenever the host VTAM application sends End Bracket (EB), Telnet sends a KRI to the client. This directly notifies the client that the keyboard can be unlocked. Without the KRI indicator, Telnet would have to make sure a WCC with the unlock keyboard flag set is sent to the client. The KRI flag is set regardless of whether the keyboard restore flag in the WCC byte is set.
  - Start Data Indicator (SDI) - When the host VTAM application sends change direction or end bracket, Telnet sends the SDI to the client. This allows the client to know exactly when data can be sent to Telnet.

  - BID - A BID sent from the host VTAM application is forwarded to the client instead of being intercepted and handled by Telnet. This allows the client to manage the BID process itself.
  - Signal Indicator - A signal that is received from the host VTAM application is forwarded to the client. When the client responds to the signal, Telnet sends a change direction indicator to the host VTAM application.
- SNA Sense Support - Allows the client to include SNA sense codes in a response message. The client retains the option of letting Telnet map the errors to an appropriate sense code by not turning on the SNA-Sense-Code indicator in the response message.

TN3270

TN3270 connections support full-screen 3270 emulation. TN3270 connections do not support:

- Device name or pool name specification
- Printers
- Client involvement with responses, SysReq, Start Data Indicator, BID, Signal or SNA Sense data.

RFC1646 defines device name specification and printer support for TN3270 connections. However, this RFC is not supported on the TN3270 Telnet server. If either of these requests is received on a TN3270 connection, the server will drop the connection.

Linemode

In some cases, the client or the application does not support full-screen presentation, or the user needs to work in a linemode environment. For these reasons, most emulators support linemode. Linemode supports a *go-ahead* function to simulate a half-duplex format. With go-ahead negotiated, the partner cannot send data until it receives a go-ahead from the current sender of data. In most cases, sessions are naturally half-duplex and the go-ahead adds unneeded transmissions. Therefore, the Telnet default is to Suppress Go Ahead (SGA). If go-ahead is needed to maintain a half-duplex format, use the NOSGA parameter. SGA or NOSGA can be coded at all three parameter block levels for different levels of granularity.

Telnet supports the following types of Linemode connections:

- Standard
- Binary
- Transform

Standard Linemode is assumed if neither DBCS transform nor BINARY linemode parameters are specified, or if the device type is not supported by transform. Standard Linemode is the only connection mode that requires translation by Telnet. Telnet provides multicultural support for standard Linemode connections. ASCII and EBCDIC code pages are the basis for translation. Telnet uses the ICONV services available in the C runtime library. For custom code page information, see the ICONV services in z/OS XL C/C++ Programming Guide. When ASCII and EBCDIC code pages are specified, a conversion descriptor will be given to Telnet. Telnet creates ASCII-EBCDIC and EBCDIC-ASCII translation tables based on the conversion descriptor. The CODEPAGE parameter is used to specify the code page names. For example:

```
CODEPAGE ISO8859-1 IBM1047
```

The possible results from CODEPAGE processing are:

- If a conversion descriptor is not returned, CODEPAGE is not coded, or there is an error in the syntax, a default code page of ISO8859-1 will be used for ASCII, and the language environment code page taken from locale information will be used as the EBCDIC code page.
- If a conversion descriptor is not returned again, a default code page of IBM-1047 will be used for EBCDIC.

- If a conversion descriptor is not returned again, predefined translation tables within Telnet will be used. These tables are similar, but not exactly the same as the tables which would have been generated if ISO8859-1 and IBM-1047 had worked. Some of the differences are noted below:

```
EBCDIC           ASCII
x'0D25' -----> x'0D0085' using ISO8859-1/IBM-1047
x'0D25' -----> x'0D0A'   using internal tables
x'15'   <----  x'0A'     using ISO8859-1/IBM-1047
x'25'   <----  x'0A'     using internal tables
```

No message is issued to the console if the first conversion succeeds. If there is any conversion failure a message is issued. If one of the later conversions succeeds, a message is issued indicating success.

If your Linemode connection does not perform correctly, the default translation tables may be causing the problem. Try the internal Telnet translation tables by specifying TNSTD for both ASCII and EBCDIC choices. For example:

```
CodePage TNSTD TNSTD
```

The internal code pages must be used together. If only one of the two internal tables is specified, then the other internal table will also be used.

CODEPAGE can be coded at all three parameter block levels for different levels of granularity.

Binary Linemode is set using the BINARYLINEMODE parameter. It indicates that Telnet should not do translation. The ASCII data from the client should be passed as-is to the VTAM application. BINARYLINEMODE or NOBINARYLINEMODE can be coded at all three parameter block levels for different levels of granularity.

Transform Linemode is set using the DBCSTRANSFORM parameter. When coded, all data that passes through Telnet will be transformed from DBCS or SBCS ASCII full screen to 3270 full screen for all supported device types. If the device type is not supported, Standard or Binary Linemode is used. DBCSTRANSFORM can be coded in TELNETPARMS or PARMSGROUP for different levels of granularity. It cannot be coded in TELNETGLOBALS. A unique logmode for transform can be set using TELNETDEVICE with a device type of TRANSFORM. Any logmode used must not support extended graphics.

Note: Transform can be used by only one port when multiple ports are active on one TCP/IP stack. DBCSTRANSFORM supports a maximum of 250 concurrent connections.

DBCSTRANSFORM can be used for either the VT100 single-byte character set (SBCS) or VT282 double-byte character set (DBCS) transform mode. When DBCSTRANSFORM is specified and the TCP/IP procedure JCL has been modified as shown below, ASCII-based terminal emulators (VT100 or VT282) will appear as full-screen 3270 terminals. Telnet receives ASCII data from the client and transforms it into SBCS or DBCS EBCDIC data, depending on the terminal type. Telnet adds appropriate SNA control bytes to give the appearance that the data is coming from a 3270 terminal. Telnet receives EBCDIC data from the host application and transforms the SNA control bytes and data into appropriate ASCII control bytes and data. The data is sent to the ASCII-based terminal where it is displayed in 3270 full-screen emulation. DBCSTRANSFORM requires additional special Data Definition (DD) statements in the TCP/IP procedure.

You must add the following three DD statements to the TCP/IP procedure JCL to support Transform:

```
//TNDBCSCN DD DSN=TCPIP.SEZAINST(TNDBCSCN),DISP=SHR
//TNDBCSXL DD DSN=TCPIP.SEZAXLD2,DISP=SHR
//TNDBCSER DD SYSOUT=*
```

- The TNDBCSCN DD statement must point to the configuration data set for 3270 DBCS transform mode. This configuration data set specifies the default DBCS conversion mode that will take effect at initialization time. Specify the CODEKIND and CHARMODE parameters according to the required DBCS code page. If CODEKIND and CHARMODE are not specified, or if the TNDBCSCN DD statement is not added, by default, CODEKIND is SJISKANJI and CHARMODE is ALPHABET. A sample can be found in SEZAINST(TNDBCSCN).
- The TNDBCSXL DD statement must point to the data set containing binary translation table code files for 3270 DBCS transform mode. The installation data set, SEZAXLD2, contains the default

binary translation table code files. The binary translation table code files for 3270 Transform can be customized by using the CONVXLAT command. See z/OS Communications Server: IP Configuration Reference for more information about customizing translation table code files. If the TNDBCSXL DD statement is not added, the following message will appear and an abend will occur:

```
IEC130I PASCAL01 DD STATEMENT MISSING
```

- The TNDBCSER DD statement defines where Transform-specific error messages are recorded. This DD statement can specify an output data set or SYSOUT=*. If the TNDBCSER DD statement is not added, transform initialization will fail.

Specifying the DBCSTRACE parameter sends detailed trace output from 3270 Transform to the location specified in the SYSPRINT output DD statement. Additional detailed trace output is also sent to TNDBCSER. Both data sets will contain detailed trace data. DBCSTRACE or NODBCSTRACE can be coded in TELNETPARMS or PARMSGROUP for different levels of granularity. They cannot be coded in TELNETGLOBALS.

Connection security

This topic describes data overrun security, Transport Layer Security (TLS), and Network Access Control.

Data overrun security

Use the MAXRECEIVE, MAXREQSESS, MAXRUCHAIN, MAXTCPSENDQ, and MAXVTAMSENDQ parameters to protect against data overrun. These parameters can be coded at all three parameter block levels for different levels of granularity.

MAXRECEIVE

This parameter limits the number of bytes received from a client without an End Of Record (EOR) being received. If the data received exceeds the limit, the connection is dropped. This parameter protects against a client stuck in a send-data loop. In general, large file transfers will not be affected because the sending client typically divides the file into smaller records that are sent. The receiving application rebuilds the file as the smaller records are received.

MAXREQSESS

This parameter limits the number of session requests received by Telnet in a 10-second period. For this parameter, a BIND received by Telnet defines a session request. If the number of BINDs received in a 10-second period exceeds the limit, an error is reported. This parameter protects against session logon loops that are possibly created by an automatic CLSDST-PASS to an inactive session. This parameter cannot protect against logon loops caused by an inactive default application and a client using auto-reconnect.

MAXRUCHAIN

This parameter limits the number of chained RUs that can be received over a session from a host application without a corresponding end chain RU. If the number of RUs exceeds the limit, the session is dropped. This parameter protects against using up large amounts of storage to hold data that is destined for a client in session with the host application.

MAXTCPSENDQ

This parameter limits the number of bytes that are queued to be sent to a Telnet client. If the queue size exceeds the limit, the connection is dropped. This parameter protects against using up large amounts of storage to hold data that is destined for an unresponsive Telnet client.

MAXVTAMSENDQ

This parameter limits the number of data segments (RPLs) queued to be sent to VTAM. If the queue size exceeds the limit, the connection is dropped. This parameter protects against using up large amounts of storage to hold data destined for a host application that is not receiving data.

Auto-reconnect loop

Without MSG07 coded a client connection error causes Telnet to drop the connection. The error may be an inactive DEFAULTAPPL or an LU assignment error. If the client has AUTO-RECONNECT specified, a continuous loop of attempts occurs. The best protection against this is to code the MSG07 parameter which keeps the client from being disconnected. However, other applications can be chosen from the error screen returned to the user. To block users from other applications, use the DEFONLY parameter.

Transport Layer Security

Note: References to RACF® apply to any SAF-compliant security product that contains the required support.

The TN3270E Telnet server (Telnet) provides the ability to secure Telnet connections with Transport Layer Security (TLS) or the Secure Sockets Layer(SSL) protocol using Application Transparent Transport Layer Security (AT-TLS) in TCP/IP. A port using AT-TLS security configuration is referred to as a TTLSPORT port. A basic port is one that does not use the TLS protocol. Connections are either secure or basic. The flows between Telnet and VTAM are unchanged.

The expired Internet Engineering Task Force (IETF) *TLS-based Telnet security* draft is supported in Telnet. This draft allows a Telnet negotiation to determine whether the client wants or supports TLS protocol prior to beginning the secure handshake. The default action that Telnet takes for a secure port is to first attempt a TLS handshake. If the client does not start the handshake within the specified handshake timeout time, an attempt is made to negotiate TLS as defined by the expired TLS-based Telnet security draft. If the client responds that it wants a secure connection, the handshake is started; if the client rejects the TLS negotiation request, the connection is closed. In this way, installations can support both types of secure clients without knowing which protocol the client is using. The default action can be changed by specifying the CONNTYPE statement described later in this topic. You can also use the CONNTYPE statement to support secure and basic connections on the same port.

Telnet server authentication and client authentication are described in Appendix B, “TLS/SSL security,” on page 1329. The Telnet server supports level 1, level 2 and level 3 client authentication. Client authentication is done with the ClientAuthType parameter of the TTLSEnvironmentAdvancedParms statement in AT-TLS policy. Level 2 and level 3 client authentication use RACF services to translate the client certificate to an associated user ID. That user ID can also be used as a client identifier.

Telnet Transport Layer Security setup

The TTLSPORT statement in the TELNETPARMS block is required to define a port as a secure port that is using AT-TLS to secure the connections.

The CONNTYPE statement is an optional statement on secure ports that provides more control over how connections initiate the TLS handshake, whether or not the connection is secure, and whether the connection is available for use. Valid CONNTYPE statement options are as follows:

- SECURE

  Indicates that the TLS handshake is used to start the connection. If the client does not start the handshake within the time specified by the handshake timeout time, an attempt is made to perform a negotiated TLS handshake (as defined by the expired IETF TLS-based Telnet security draft). If the client rejects TLS, the connection is closed.

- NEGTSECURE

  Indicates that the client supports the expired IETF TLS-based Telnet security draft. A Telnet negotiation with the client determines whether the client is willing to enter into a secure connection. If the client agrees, a TLS handshake is started and secure protocols are used for all subsequent communication. If the client rejects TLS, the connection is closed. You should consider using this option only if you know that the Telnet secure clients connecting into the port are all using the protocol defined by the expired TLS-based Telnet security draft. With this option, the TLS handshake is not attempted until a positive response to the Telnet DO_StartTLS IAC is received. This avoids the timeout delay that can occur when a TLS handshake is immediately started (as occurs with the CONNTYPE SECURE option), but the client is expecting the protocol used by the expired TLS-based Telnet security draft. Use the SECURE option

instead of the NEGTSECURE option in case some clients in your network do not support the expired TLS-based Telnet security draft.

- BASIC

  Indicates that a basic connection is established.

- ANY

  Indicates that the connection can be either secure or basic. Telnet first tries a standard TLS handshake. If the handshake times out, a negotiated TLS connection (see the CONNTYPE NEGTSECURE option description) is attempted:

  – If the client is willing to enter into a secure connection, secure protocols are used for all subsequent communication.
  – If the client is not willing to enter into a secure connection, a basic connection is established.

- NONE

  Indicates that no connection is allowed and the connection will be closed. If this option is specified in the TELNETPARMS block, a PARMSMAP statement must cover every allowable connection, and the related PARMSGROUP statement must specify the connection type on the CONNTYPE statement.

If the CONNTYPE statement is not specified, by default, secure ports are CONNTYPE SECURE and basic ports are CONNTYPE BASIC.

Using one port for both basic and secure connections

You can use the CONNTYPE statement to modify connection types on a single port. Allowing a port to support both basic and secure connections assumes that either of the following statements are true:

- The installation allows the client to determine the connection type.
- A subset of the connections that should use a particular connection security type can be identified by Client Identifier.

In the first case, specify CONNTYPE ANY. If the port was defined as a secure port but the client wants a basic connection, there is a slight delay before connection negotiation begins. This is because when CONNTYPE ANY is coded, Telnet first attempts a TLS handshake to ensure that the client is not requesting TLS support. It is only after the handshake times out and negotiated security is rejected that the basic connection negotiation begins.

In the second case, the TELNETPARMS block should specify the default connection security type (see the CONNTYPE statement). For connections with different connection security requirements:

- Identify the clients by Client Identifier.
- Create a group using the PARMSGROUP statement with the alternate CONNTYPE definitions.
- Map the group created with the PARMSGROUP statement to the clients using the PARMSMAP statement.

Configuring Telnet security using AT-TLS

The TTLSPORT statement in the TELNETPARMS block indicates that the port uses AT-TLS to manage System SSL. All TTLSPORT ports must be defined by specifying a TELNETPARMS block for each port.

Other than the CONNTYPE statement, all security configuration is done in AT-TLS policy. For details about AT-TLS setup, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121. For Policy Agent setup and AT-TLS policy statements, see z/OS Communications Server: IP Configuration Reference. A sample list of tasks to perform for AT-TLS policy includes:

1. Be sure that the TCP/IP stack profile includes the TCPCONFIG statement with the TTLS parameter.
2. Permit Policy Agent and any other required administrative application to the RACF resource EZB.INITSTACK.*sysname*.*tcpname* in the SERVAUTH class.
3. Define the pagent environment file on the STDENV DD statement in Policy Agent JCL. For example:

```
//STDENV   DD PATH='/etc/pagent/pagent.env',PATHOPTS=(ORDONLY)
```

4. In the pagent environment file, point to a configuration file. For example:

```
PAGENT_CONFIG_FILE=//'SYS1.TCPPARMS(PAGENT)'
```

5. In the configuration file, set up policy files for each TCP/IP stack image. For example:

```
TcpImage TCP1 /etc/pagent/TCP1.policy FLUSH
TcpImage TCP2 /etc/pagent/TCP2.policy FLUSH
```

6. In the TcpImage file, point to the TTLS configuration file. For example:

```
TTLSConfig /etc/pagent/pagttls1.ttls
```

7. In the TTLS configuration file, code the TTLSRule, TTLSGroupAction, TTLSEnvironment, and TTLSConnectionAction statements. Be sure to set the ApplicationControlled parameter to the value On in the TTLSEnvironmentAdvancedParms statement. For example:

```
TTLSRule        tn_serv1
{
  LocalPortRange 23
  Direction      Inbound
  Jobname TCP1
  TTLSGroupActionRef tn_grp_act
  TTLSEnvironmentActionRef tn_env_act
}

TTLSGroupAction tn_grp_act
{
 TTLSEnabled On
 Trace  7
 GroupUserInstance  1
}

TTLSEnvironmentAction tn_env_act
{
  HandshakeRole Server
  TTLSKeyringParms
  {
     Keyring TN3270E/TNsafkeyring
  }
  TTLSEnvironmentAdvancedParms
  {
   ApplicationControlled On
  }
 EnvironmentUserInstance  1
}
```

8. Verify that the policy is correctly entered by using the z/OS UNIX **pasearch** command to query information from the z/OS UNIX Policy Agent. Issue the **pasearch -t** command from the z/OS UNIX System Services shell. If you have multiple TCP/IP stacks that are active, issue the **pasearch -t -p** *procname* command to query a specific TCP/IP stack. The **pasearch** command is a Policy API (PAPI) application. If you have never run a PAPI application, you might receive a message indicating that the papi.dll file was not found. For more information about PAPI and running PAPI applications, see z/OS Communications Server: IP Programmer's Guide and Reference.

Telnet profile example

This example defines three ports with the following characteristics:

- Port 23 allows only basic connections.
- Ports 992 and 1023 are enabled for secure connections defined by AT-TLS policy.
- Port 992 allows only secure connections. No client authentication is requested.
- Port 1023 allows both basic and secure connections. The installation wants the following characteristics for port 1023:

- The system administrator is at IP address 10.1.3.3 and wants the capability to choose to connect with secure or basic connections.
- Buildings A and B are local and do not need connection security. The clients in these buildings have identifiable subnetworks (10.1.1.0/24 and 10.1.2.0/24, respectively). The installation wants these clients to use basic connections to avoid the encryption overhead.
- TLS security with client authentication is required for all other connections.

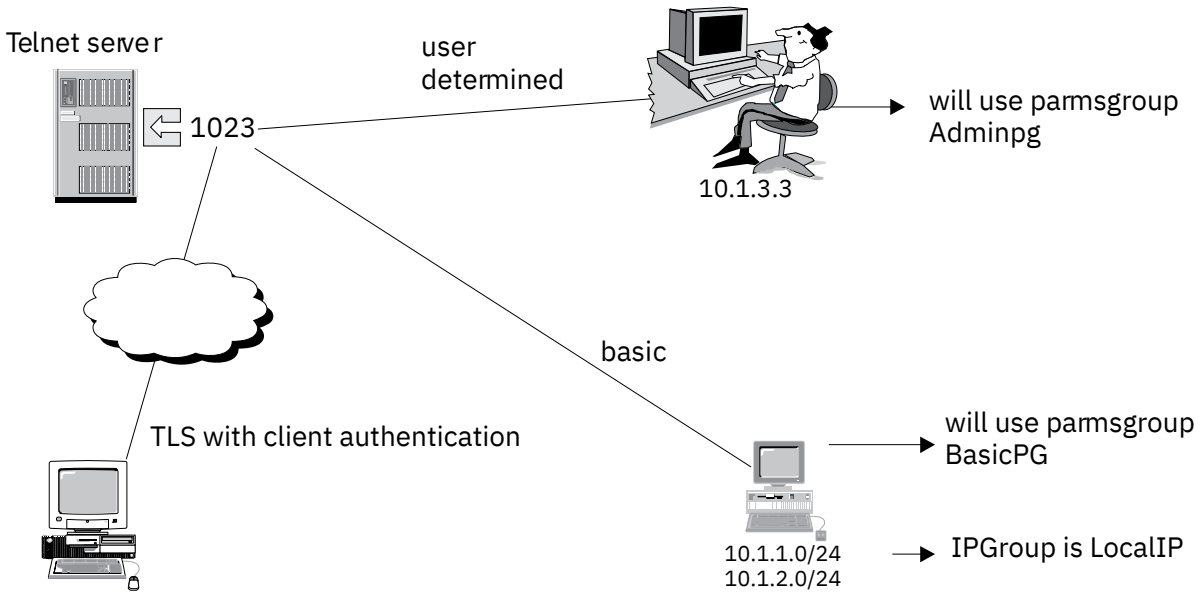


Figure 87. Port 1023 connection characteristics

Note: Definitions that are applicable to TLS connection security are the only definitions shown; additional parameters might be needed. Assume that all connections go through TCP/IP stack with job name TCP1.

TCP/IP configuration statements:

```
⋮
 TCPCONFIG TTLS
⋮
```

Telnet profile statements:

```
TELNETPARMS             ; basic port does not support secure connections
 Port 23
ENDTELNETPARMS

TELNETPARMS             ; port that allows only secure connections
 TTLSPORT 992           ; no client authentication requested
ENDTELNETPARMS

TELNETPARMS             ; port that allows secure and BASIC connections.
 TTLSPORT 1023          ; note: BEGINVTAM block has PARMSGROUP that may override
 CONNTYPE SECURE        ; this CONNTYPE setting. If not, SECURE will be default.
ENDTELNETPARMS

BEGINVTAM
 Port 1023
 ...                    ; Mapping statements
 IPGROUP LocalIP        ; Subnets for buildings A and B
   255.255.255.0:10.1.1.0
   255.255.255.0:10.1.2.0
 ENDIPGROUP
 PARMSGROUP BasicPG     ; override default ConnType
   CONNTYPE BASIC       ; support basic connections if mapped to this group
 ENDPARMSGROUP
 PARMSGROUP AdminPG     ; override default ConnType
   CONNTYPE ANY         ; allow any type of connections if mapped to this group
 ENDPARMSGROUP
 PARMSMAP AdminPG 10.1.3.3 ; this ip address can use secure or basic connections
 PARMSMAP BasicPG localIP  ; hosts defined in IPGROUP localIP will use basic
                           ; connections as defined in PARMSGROUP BasicPG
ENDVTAM

BEGINVTAM
 Port 992 23
 ...                    ; Mapping statements
                        ; no PARMSGROUP defined for these ports
                        ; TELNETPARMS definitions used for all connections
ENDVTAM
```

AT-TLS policy statements:

```
TTLSRule tn992_serv
{
  LocalPortRange 992
  Direction Inbound
  Jobname TN3270A
  TTLSGroupActionRef tn_grp_act
  TTLSEnvironmentActionRef tn992_env_act
}

TTLSRule tn1023_serv
{
  LocalPortRange 1023
  Direction Inbound
  Jobname TN3270A
  TTLSGroupActionRef tn_grp_act
  TTLSEnvironmentActionRef tn1023_env_act
}

TTLSGroupAction tn_grp_act
{
  TTLSEnabled On
  Trace 7
```

```
  GroupUserInstance 1
}

TTLSEnvironmentAction tn992_env_act
{
  HandshakeRole Server
  TTLSKeyringParmsRef tn_keyring
  TTLSEnvironmentAdvancedParms
  {
    ApplicationControlled On
  }
  EnvironmentUserInstance 1
}

TTLSEnvironmentAction tn1023_env_act
{
  HandshakeRole ServerWithClientAuth
  TTLSKeyringParmsRef tn_keyring
  TTLSEnvironmentAdvancedParms
  {
    ClientAuthType Required
    ApplicationControlled On
  }
  EnvironmentUserInstance 1
}

TTLSKeyringParms tn_keyring
{
  Keyring TN3270E/TNsafkeyring
}
```

Network Access Control

Network Access Control (NAC) limits user access to certain IP security zones defined by the NETACCESS statement. A security product, such as RACF, is used to check the permission of user IDs to send data to or receive data from these security zones. The NAC user ID is based on the Telnet address space user ID information.

The NACUSERID parameter provides more control over Network Access Control checking for Telnet. This parameter is used to associate Telnet ports with a specified user ID that is defined to the security server. The user ID specified on the NACUSERID parameter must be a valid user ID defined to the security server. If not, the Telnet port will fail initialization. NACUSERID can be coded in TELNETGLOBALS to affect all ports or TELNETPARMS to affect a single port. NACUSERID cannot be coded in PARMSGROUP. Specify NONACUSERID to disable a higher level specification. For example, a TN3270E Telnet server with an address space user ID of user1 can specify in TELNETGLOBALS the statement NACUSERID user2. If one port should instead be controlled by user1, the TELNETPARMS statement for that port should be NONACUSERID to disable the user2 specification in TELNETGLOBALS.

When Telnet is modified with a VARY TCPIP,*tnproc*,OBEYFILE command, the NACUSERIDs are reverified for the Telnet ports defined in the data set referenced by the command. If a Telnet port has NACUSERID *NAC_name_1*, you cannot use the VARY TCPIP,*tnproc*,OBEYFILE command to change that port's NACUSERID to *NAC_name_2*. The port must first be stopped, and then started with the new *NAC_name_2* value using the VARY TCPIP,*tnproc*,OBEYFILE command.

The NETACCESS statement in the TCP/IP profile is used to configure portions of your IP network into named security zones. Each defined security zone must have a SERVAUTH profile for the resource named EZB.NETACCESS.*sysname*.*tcpname*.*zonename*. The user ID associated with the Telnet port must have READ access to the security zone that maps its bind address (0.0.0.0/32 for INADDR_ANY or ::/128 for the IPv6 unspecified address, in6addr_any, unless overridden by the PORT statement in the TCP/IP profile) and to every security zone that maps client IP addresses that Telnet is to accept connections from on this port.

For more information, see “Network access control” on page 140.

Connection persistence

Several timers are available in Telnet to control how long connections stay up. The list includes:

- INACTIVE - How long a terminal connection can be idle with no SNA data traffic before the connection is dropped.
- PRTINACTIVE - How long a printer connection can be idle with no SNA data traffic before the connection is dropped.
- PROFILEINACTIVE - How long a Telnet connection can be active with no active SNA session, while it is associated with a Telnet profile that is not the current profile.
- KEEPINACTIVE - How long a KEEPOPEN connection can be idle with no SNA session before the connection is dropped. When a KEEPOPEN connection is in session with a SNA application the INACTIVE timer is used instead of the KEEPINACTIVE timer.
- SCANINTERVAL - How often Telnet runs the list of connections looking for potentially lost connections. Because of the methodology, it also determines how long Telnet will wait for a TIMEMARK response before assuming the connection is lost.
- TIMEMARK - How long a connection is active without receiving any data before Telnet sends a TIMEMARK command which acts as an "are you there".

To facilitate these timers, Telnet records the time at which data is received from the client, received from VTAM, or sent to VTAM. Data received from the client is used by SCANINTERVAL/TIMEMARK to measure idle time on the connection. Data received from or sent to VTAM is used by the INACTIVE family of timers to measure idle time without SNA data traffic.

The INACTIVE family of timers

INACTIVE, PRTINACTIVE, PROFILEINACTIVE, and KEEPINACTIVE all share one timer associated with a port profile to reduce system overhead. The timer with the smallest value defined in TELNETGLOBALS, TELNETPARMS, or PARMSGROUP for that port profile is used to define how often the connections are checked.

For example, assume KEEPINACTIVE is defined as 1800, PROFILEINACTIVE is 1800 by default, INACTIVE is defined as 3000, and PRTINACTIVE is defined as 5400 in a profile. The Telnet timer will run every 1800 seconds. Therefore, every time the timer expires, Telnet will check each KEEPOPEN connection not in session to see if there has been a SNA session created in the prior 1800 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-K. For PROFILEINACTIVE, Telnet checks each connection associated with a profile that is not current to determine whether a SNA session existed in the previous 1800 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-PF. Telnet will also check each terminal connection to see if there has been any SNA data traffic in the prior 3000 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-S. Telnet also will check each printer connection to see if there has been any SNA data traffic in the prior 5400 seconds. If not, the connection is dropped with DEBUG SUMMARY message CONN DROP reason INACT-P.

Setting KEEPOPEN to the smallest time was done as an example. Any of the four timers could be the smallest. Also, because all inactivity checks are done by one timer, a connection check could occur just before the connection would be considered timed out. The connection will remain active until the next check is made.

Using the example above, if a terminal connection has had no SNA activity in the past 2999 seconds when a check is made, the connection is not dropped. Another check is done 1800 seconds later and the connection is dropped, but the connection will have remained active for 4799 seconds instead of the specified 3000 seconds.

SCANINTERVAL and TIMEMARK

SCANINTERVAL and TIMEMARK are used together to determine if a connection has been lost. These parameters can be specified in TELNETGLOBALS, TELNETPARMS, and PARMSGROUP. The smallest SCANINTERVAL value is used to define how often the connections are checked. If the SCANINTERVAL value is greater than the TIMEMARK value, the SCANINTERVAL value is reset to the TIMEMARK value. Whenever data is received from the client, Telnet records the time. Telnet checks all connections at regular intervals defined by the SCANINTERVAL value. Each connection is checked to see if any data has been received from the client in the past TIMEMARK period of time. If not, a TIMEMARK command is

sent to the client which acts as an "are you there" and Telnet remembers a TIMEMARK was sent to this client. During the next check at SCANINTERVAL time later, each connection is again checked to see if any data has been received from the client. If not, and a TIMEMARK was sent on the previous check, the connection is dropped with DEBUG SUMMARY message CONN DROP reason TIMEMARK.

For example, assume the values for SCANINTERVAL and TIMEMARK are 1800 and 10800, respectively. That means every 30 minutes all connections are checked to see if any data has been received in the last 3 hours. If not, a TIMEMARK is sent to the client. 30 minutes later Telnet checks the connections again. If the client responded to the TIMEMARK or sent in actual data of some type Telnet leaves the connection active. If nothing has been received Telnet drops the connection.

A SCANINTERVAL check could occur just before the last data received is old enough to trigger sending a TIMEMARK. The connection remains active until the next SCANINTERVAL is made. Then a TIMEMARK is sent, and at the next SCANINTERVAL the connection is dropped.

Using the example above, SCANINTERVAL checks a connection that received data 2 hours, 59 minutes ago. No TIMEMARK is sent. The next SCANINTERVAL runs 30 minutes later. Now the data received time is greater than 3 hours and a TIMEMARK is sent. At the next SCANINTERVAL, the connection is dropped. The connection's last activity was 3 hours 59 minutes ago.

Tip: Use Scaninterval and Timemark to find abandoned connections that do not require quick reset. Scaninterval and Timemark are intended to eventually clean up abandoned connections. They should not be used as an immediate reset function. If immediate reset of lost connections is needed, use the CheckClientConn parameter.

Setting the timers

Caution must be used in setting these timers. Setting the INACTIVE family of timers or SCANINTERVAL timer too low could cause excessive CPU usage. Setting the TIMEMARK value too low could also cause excessive flooding of the network with TIMEMARK commands or high storage usage. For example, these timers should take into account extended breaks such as lunch. If TIMEMARK is smaller than the lunch break time, the network may be flooded with TIMEMARK commands around the lunch hour. Be aware of the default values and be sure to set appropriate values for the situation.

MSG07 and LUSESSIONPEND

MSG07 and LUSESSIONPEND are Telnet parameter statements that define what Telnet should do in case of a session setup error and after normal logoff when the client is emulating a terminal. These parameters do not affect a printer connection.

- Connection negotiation error - If any problems occur during negotiation nothing can be done to keep the connection. If appropriate, Telnet will send the client an error code to help inform the client why the connection was dropped and issue a CONN DROP DEBUG message at the console.
- Session setup error - If a problem occurs during session setup such as an application name that is not valid, session request failure, or a BIND error, Telnet will drop the connection and issue a CONN DROP DEBUG message. The user cannot get to any application other than the default. No error messages are sent to the user and auto-reconnect loops are possible. For these reasons it is recommended that MSG07 always be used. If the MSG07 parameter is coded, the connection will not be dropped and an error message will be sent to the user. MSG07 function applies to any connection mode whether or not USS tables are mapped to the client. If a USS table is used, the user can press the CLEAR key to return to the USSMSG10 screen. If the LUMAP-DEFAPPL or PRTMAP-DEFAPPL statement is coded and the default application is not available, an error screen will be sent to the client whether or not MSG07 is coded.
- Normal Session Logoff - When the user logs off a session using a normal logoff, Telnet drops the connection. If the user typically logs on to another application after logging off the first application, it might be more efficient to present the user with another solicitor (or USSMSG10) screen or for Telnet to initiate a new session with the default application after logoff. This can be accomplished by coding the LUSESSIONPEND parameter. Code LUSESSIONPEND to go through the initial database lookup again after session logoff. Later results will be identical to the first lookup. If a default application for the client exists, Telnet will immediately initiate another session request; otherwise, a USSMSG10 screen or

solicitor screen is sent to the user. When LUSESSIONPEND is coded, the connection remains active but terminal LU ACBs are closed.

- SYSREQ LOGOFF - When the user logs off of a session using a SYSREQ LOGOFF sequence (TN3270E connection supported) and LUSESSIONPEND is coded, Telnet does not drop the connection. Instead, the user is presented with a solicitor (or USSMSG10) screen. If DEFAULTAPPL is in effect, Telnet again requests a session with the default application.
- USS LOGOFF - When the user issues a LOGOFF command from the USSMSG10 screen, the connection is dropped whether or not the LUSESSIONPEND parameter is coded.

Mapping Objects to Client Identifiers

Telnet provides flexibility for mapping Objects to clients based on Client Identifiers. This topic provides definitions, rules, and examples of many mapping methods. Examples start with simple concepts, then progress to more complicated concepts showing interaction between mapping statements. All mapping statements are specified in the BEGINVTAM block. See z/OS Communications Server: IP Configuration Reference for statement rules not discussed here.

The general relationship of mapping statements is:

MAP OBJECTS to clients based on *CLIENT IDENTIFIER*

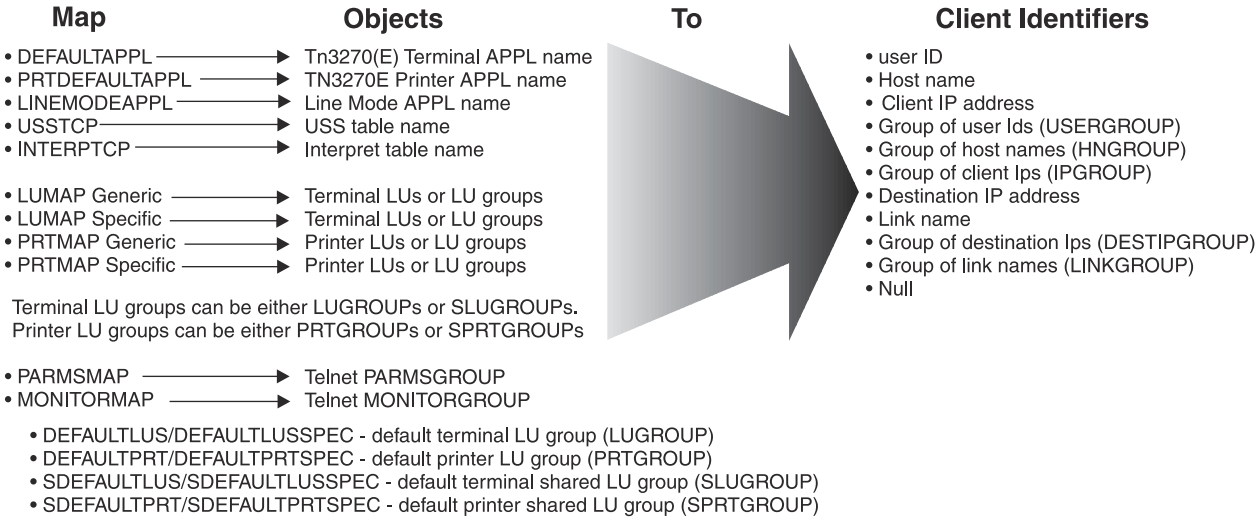


Figure 88. Mapping model

Telnet tries to assign all 11 Objects to a client based on the mapping statements when the connection is accepted. The search for Objects continues until all Objects are found or until all mapping statements are checked.

Objects

When a client connection request is made, Telnet must assign an LU name to represent the client. Optionally, a USS table, default application, unique parameters defined in the PARMSGROUP statement, or network monitoring can be assigned to the connection. See “Mapping Objects to Client Identifiers” on page 607 for details about how these objects are mapped to clients. The complete list of objects follows:

- TN3270 or TN3270E terminal application name – The DEFAULTAPPL mapping statement maps the TN3270 or TN3270E terminal application Object to a terminal client. When a TN3270 or TN3270E connection is negotiated, Telnet immediately initiates a session request to the VTAM application.
- TN3270E printer application name – The PRTDEFAULTAPPL mapping statement maps the TN3270E printer application Object to a printer client. When a TN3270E printer connection is negotiated, Telnet immediately initiates a session request to the VTAM application.

- Line Mode application name – The LINEMODEAPPL mapping statement maps the linemode application Object to a client. When a linemode connection is negotiated, Telnet will immediately initiate a session request to the VTAM application.
- USS table name – The USSTCP mapping statement maps the USS table Object to a client. When a TN3270 or a TN3270E connection is negotiated, Telnet will send a USSMSG10 screen to the client. A special case condition exists when an application name and a USS table are both mapped to the client by the exact same Client Identifier. In this case, Telnet will immediately initiate a session request to the VTAM application and use the USS table for error messages.
- Interpret table name – The INTERPTCP mapping statement maps the Interpret table Object to a client. When a TN3270 or a TN3270E connection is negotiated, Telnet uses the Interpret table to modify USS commands. The client must have a USS table mapped to it for the Interpret table to be used.
- Terminal LU, local LU group (LUGROUP), or shared LU group (SLUGROUP) (Generic) - The Generic LUMAP mapping statement maps a single LU, LUGROUP, or SLUGROUP Object to a client.
  - For single LU mappings, Telnet assigns the LU name to the connection if the LU is available.
  - For LUGROUP mappings, Telnet assigns an available LU from the group to the connection.
  - For SLUGROUP mappings, Telnet requests that the LUNS allocate an available LU from the group and then assigns it to the connection.

  An LU is required to represent the client when a VTAM session is initiated. DEFAULTLUS is a default local terminal LUGROUP Object mapped Generically to the NULL Client Identifier. SDEFAULTLUS is a default shared terminal SLUGROUP Object mapped Generically to the NULL Client Identifier.
- Terminal LU, LUGROUP, or SLUGROUP (Specific) - The Specific LUMAP mapping statement maps a single LU, LUGROUP, or SLUGROUP Object to a client.

  Unlike the Generic mapping in which Telnet assigns the LU, the Specific mapping requires the client to specify the LU name that it wants. Telnet verifies that the LU is mapped and available. The specified LU name can be either a mapped single LU, an LU within a mapped LUGROUP or SLUGROUP, or the group name of the mapped LUGROUP or SLUGROUP.
  - If the client specifies an LU name within a SLUGROUP, Telnet requests the LUNS to verify that the LU is available and allocate the LU to this Telnet.
  - If the client specifies an LUGROUP name, Telnet assigns an available LU from within the group.
  - If the client specifies an SLUGROUP name, Telnet requests the LUNS to allocate an available LU from the group and then assigns the LU.

  DEFAULTLUSSPEC is a default local terminal LUGROUP Object mapped Specifically to the NULL Client Identifier. SDEFAULTLUSSPEC is a default shared terminal SLUGROUP Object mapped Specifically to the NULL Client Identifier.
- Printer LU, local printer group (PRTGROUP), or shared printer group (SPRTGROUP) (Generic) - The Generic PRTMAP mapping statement maps a single LU, PRTGROUP, or SPRTGROUP Object to a client.
  - For single LU mappings, Telnet assigns the LU name to the connection if the LU is available.
  - For PRTGROUP mappings, Telnet assigns an available LU from the group to the connection.
  - For SPRTGROUP mappings, Telnet requests the LUNS to allocate an available LU from the group and then assigns it to the connection.

  An LU is required to represent the client when a VTAM session is initiated. DEFAULTPRT is a default local printer PRTGROUP Object mapped Generically to the NULL Client Identifier. SDEFAULTPRT is a default shared printer SPRTGROUP Object mapped Generically to the NULL Client Identifier.
- Printer LU, PRTGROUP, or SPRTGROUP (Specific) - The Specific PRTMAP mapping statement maps a single LU, PRTGROUP, or SPRTGROUP Object to a client.

  Unlike the Generic mapping in which Telnet assigns the LU, the Specific mapping requires the client to specify the LU name that it wants. Telnet verifies that the LU is mapped and available. The specified LU name can be either a mapped single LU, an LU within a mapped PRTGROUP or SPRTGROUP, or the group name of the mapped PRTGROUP or SPRTGROUP.

  - If the client specifies an LU name within a SPRTGROUP, Telnet requests that the LUNS verify that the LU is available and allocate the LU to this Telnet.
  - If the client specifies a PRTGROUP name, Telnet assigns an available LU from within the group.
  - If the client specifies a SPRTGROUP name, Telnet requests that the LUNS allocate an available LU from the group and then assigns the LU.

  DEFAULTPRTSPEC is a default local printer PRTGROUP Object mapped Specifically to the NULL Client Identifier. SDEFAULTPRTSPEC is a default shared printer SPRTGROUP Object mapped Specifically to the NULL Client Identifier.
- Telnet PARMSGROUP – The PARMSMAP mapping statement maps the PARMSGROUP Object to a client. The parameters in the group override parameter values specified in either TELNETGLOBALS or TELNETPARMS.
- Telnet MONITORGROUP – The MONITORMAP mapping statement maps the MONITORGROUP Object to a client. The parameters in the group define what monitoring measurements will be done for a mapped client. For details, see "Connection monitoring mapping statement" on page 649.

Rule: All LUGROUPs and all PRTGROUPs, whether local or shared, that are used on a given profile must have unique names.

The two following statements are not Objects but can affect application and LU Object usage:

- ALLOWAPPL – This statement allows client access to applications and optionally maps or confirms the mapping of an LU name to the client based on the application name chosen. DEFAULTAPPL application names are presumed allowed and do not require the ALLOWAPPL statement for Telnet acceptance. However, ALLOWAPPL may be used by default applications for LU assignment and other advanced functions.
- RESTRICTAPPL – This statement restricts Telnet acceptance of application names to only users that specify an acceptable User ID and password. It also optionally maps or confirms the mapping of an LU name to the client based on the application name and User ID chosen.

Client Identifiers

One client can be represented by many different Client Identifiers. For example, Telnet might assign an LU based on client host name, assign an application based on a client IP address, and assign a USS table based on connection link or interface name. See "Mapping Objects to Client Identifiers" on page 607 for details about how these Client Identifiers are used to map Objects. In some cases, two different Client Identifiers that represent the same client are used on mapping statements to map the same type of Object. In these cases, Telnet must determine which Client Identifier to use when assigning the Object. See "Client Identifier selection rules" on page 612 for more details. The complete list of Client Identifiers and mapping examples follow:

- User ID or USERGROUP name - AT-TLS parameter ClientAuthType SAFCheck requires the client to send a client certificate to authenticate itself to the server. That certificate must be associated with a valid z/OS user ID in the security product such as RACF.The resulting User ID is associated with the connection. Objects can be mapped to the connection based on an exact User ID, or Objects can be mapped to a USERGROUP name containing exact User IDs and wildcarded User IDs. For example, mobile employees need to be assigned a unique set of LU names and the manager must always be assigned LU name LUMOBL01. These employees are not within a secure network and always use client authenticated secure connections. Their certificates are translated to User IDs by Telnet.

```
USERGROUP   USGMOBL1
   MOBL0002 MOBL0003
   MOBL1%%C
ENDUSERGROUP
LUGROUP LUGMOBL1
   LUMOBL02..LUMOBL20
ENDLUGROUP
LUMAP  LUMOBL01  USERID,MOBL0001       ; mgr mapping
LUMAP  LUGMOBL1  USERGRP,USGMOBL1      ; employee mapping
```

Rule: The specification of the Client Identifier type USERID is required on the mapping statement. If you do not specify this type, Telnet assumes that the name is a link or interface name.

Tip: The specification of the Client Identifier type USERGRP is optional. The following statement is equivalent to the last LUMAP statement in the previous example:

```
LUMAP  LUGMOBL1  USGMOBL1
```

- Host name or HNGROUP name - If the network dynamically assigns IP addresses, the same client will not have the same IP address from one connection to the next. With static host names, Objects can be mapped to clients based on their host name, or Objects can be mapped to HNGROUP names containing exact host names and wildcarded host names. For example, LUADMNM is mapped to exact host name ADMIN.DEPT1.GROUP1.COM, and application INVENTRY is mapped to HNGROUP name HNGINV.

```
HNGROUP HNGINV
   INV1.DEPT1.GROUP1.COM
   *.DEPT3.GROUP1.COM
   **.GROUP3.COM
ENDHNGROUP
LUMAP LUADMNM  HOSTNAME,ADMIN.DEPT1.GROUP1.COM
DEFAULTAPPL    INVENTRY  HNGRP,HNGINV
```

Tip: The specification of the Client Identifier types HOSTNAME and HNGRP is optional. The following two mapping statements are equivalent to the last two statements in the previous example:

```
LUMAP LUADMNM  ADMIN.DEPT1.GROUP1.COM
DEFAULTAPPL    INVENTRY  HNGINV
```

- Client (source) IP address or IPGROUP name - Client IP address is the most common method used to map Objects to the client. In a static network, Objects can be mapped to clients based on the exact IP address, or Objects can be mapped to IPGROUP names containing exact IP addresses and subnets. For example, LUADMN is mapped to exact IP address 1.1.1.1, and application PAYROLL is mapped to IPGROUP name IPGPAY.

```
IPGROUP IPGPAY
   1.1.2.2  1.1.2.3                 ;IPv4 addresses
   255.255.0.0:2.2.0.0              ;IPv4 subnet
   2001:0DB8:9:11:15:4              ;IPv6 address
   6C11:10::0/96                    ;IPv6 subnet
   6.1.3.4..6.1.3.8                 ;IPv4 range
   2AB0::12:5:1321..2AB0::12:5:1410  ;IPv6 range
ENDIPGROUP
LUMAP LUADMN IPADDR,1.1.1.1
DEFAULTAPPL  PAYROLL IPGRP,IPGPAY
```

Tips:

– The specification of the Client Identifier types IPADDR and IPGRP is optional. The following two mapping statements are equivalent to the last two statements in the previous example:

```
LUMAP LUADMN 1.1.1.1
DEFAULTAPPL  PAYROLL IPGPAY
```

– The IP/subnet combinations of 0.0.0.0:0.0.0.0 (IPv4 only) and 0::0/0 (IPv4 and IPv6) are special cases that include all connections. This might be useful if you want to have a default mapping with a higher priority than the NULL client identifier.

– The client IP address can be either an IPv4 or IPv6 IP address. IP address ranges can also be specified and are treated as if individual IP addresses were coded. An IPv4 range can vary in the last octet only. An IPv6 range can vary in the last two hexadecimal bytes only.

- Destination IP address or DESTIPGROUP name - A destination IP address is the host address that is the destination for a Telnet connection. Linkname can be used as a Client Identifier to map Objects to destination IP addresses when the linkname is static and defined in the profile. However, if the destination IP address is a dynamic Virtual IP Address (VIPA) , the linkname is not known before the VIPA is created. In this case, destination IP address is the ideal solution. In other cases, specifying the destination IP address in the Telnet profile may be more clear than specifying the linkname. For

example, two TCP/IP stacks are backups for each other. Telnet connections to stack 1 (VIPA 5.5.5.1) use logon manager application APPL1 by default, and connections to stack 2 (VIPA 51CB:C3E4::9:4) use logon manager application APPL2 by default. If one of the stacks becomes unavailable, the other will take over and dynamically add the failing stack's VIPA. The dynamic linkname created is not easily predicted. Use the following statements in the profile of each stack to ensure users connecting to 5.5.5.1 always get APPL1 and users connecting to 51CB:C3E4::9:4 always get APPL2 regardless of which stack is used.

```
DEFAULTAPPL  APPL1 DESTIP,5.5.5.1
DEFAULTAPPL  APPL2 DESTIP,51CB:C3E4::9:4
```

Rule: The specification of the Client Identifier type DESTIP is required on the mapping statement. If you do not specify this type, Telnet assumes that the IP addresses are client (source) IP addresses.

Tip: When the destination IP address is the IP address of a dynamic XCF address, multiple linkname values can be associated with the IP address. Telnet will use the first linkname associated with the IP address in the home list. If a dynamic XCF destination is used as a Client Identifier, it is recommended that DESTIP be used instead of linkname. Results can vary using linkname.

- Linkname or LINKGROUP name - A linkname is defined by the TCP/IP LINK or INTERFACE statement. The linkname defines a host IP address that is a destination address for clients connecting to Telnet. Linkname can be useful in cases where Object assignment is dependent on the client destination IP address instead of the client source IP address. Several linknames can be defined and the same LU mapping or other Object mapping might be wanted for several linknames. In this case, a LINKGROUP can be defined and used on a single mapping statement. For example, based on the statements below, a client connecting to LINK1 IP address will be assigned an LU from the LUGROUP name LUGLNKS and will establish a session with TPX1. A client connecting to LINK2 IP address will be assigned an LU from the LUGROUP name LUGLNKS and will establish a session with TPX2. Because LINK1 and LINK2 are not group names, host names, or IP addresses, they are assumed to be linknames. The Client Identifier type, LINKNAME, can be used for clarity but is not required.

  ```
  LINKGROUP    LNKGRP1
     LINK1 LINK2
  ENDLINKGROUP
  LUMAP        LUGLNKS  LINKGRP,LNKGRP1
  DEFAULTAPPL  TPX1     LINKNAME,LINK1
  DEFAULTAPPL  TPX2     LINKNAME,LINK2
  ```

  Tips:

  – The specification of the Client Identifier types LINKNAME and LINKGRP is optional. The following three mapping statements are equivalent to the last three statements in the previous example:

    ```
    LUMAP        LUGLNKS  LNKGRP1
    DEFAULTAPPL  TPX1     LINK1
    DEFAULTAPPL  TPX2     LINK2
    ```

  – When the destination IP address is the IP address of a dynamic XCF address, multiple linkname values can be associated with the IP address. Telnet will use the first linkname associated with the IP address in the home list. If a dynamic XCF destination is used as a Client Identifier, it is recommended that DESTIP be used instead of linkname. Results can vary using linkname.

- NULL (no Client Identifier) - The NULL Client Identifier type indicates that no Client Identifier was specified. The NULL Client Identifier is valid on the DEFAULTAPPL, LINEMODEAPPL, USSTCP, and INTERPTCP mapping statements. It is the implied Client Identifier for the DEFAULTLUS, DEFAULTLUSSPEC, DEFAULTPRT, and DEFAULTPRTSPEC Objects. ParmsGroup and MonitorGroup are the only Objects that cannot be mapped to the NULL Client Identifier. The NULL Client Identifier mapped Objects are the last Objects checked when assigning Objects to a client. For example, assume a client does not match any Client Identifier in the profile for DEFAULTAPPL or USSTCP. You can put the user into session with a security application, named SecAppl, that can verify the user is authorized to use the company's system. The Client Identifier field is blank.

  ```
  DEFAULTAPPL SECAPPL
  ```

Client Identifier selection rules

When Client Identifiers are used together, conflicts might occur. For example, host name NAME1.HOST1.COM may also be IP address 1.2.3.4. If the following DEFAULTAPPL statements exist, only one of the applications can be chosen.

```
DEFAULTAPPL  TSO  NAME1.HOST1.COM
DEFAULTAPPL  CICS 1.2.3.4
```

If USSTCP and DEFAULTAPPL have the same Client Identifier, DEFAULTAPPL will be used. For detailed information, see "Resolving DEFAULTAPPL and USS table conflicts" on page 622.

Telnet uses a very specific Client Identifier hierarchy when assigning Objects, as shown in "The mapping rule search order" on page 612

The mapping rule search order

- Exact client identifier:
  - 1) User ID, 2) hostname, 3) IP address
- Exact client identifier in a group definition:
  - 4) User group, 5) hostname group, 6) IP address group
- Wildcard match for client identifier in a group definition:
  - 7) User group, 8) hostname group, 9) IP address group
- Exact destination:
  - 10) destination IP address, 11) link or interface name
- Exact destination in a group definition:
  - 12) destination IP address group, 13) link or interface name group
- Wild card match for destination in a group definition:
  - 14) destination IP address group, 15) link or interface name group
- Null client ID
  - 16) DEFAULTAPPL, LINEMODEAPPL, USSTCP, INTERPTCP, DEFAULTLUS, DEFAULTLUSSPEC, DEFAULTPRT, DEFAULTPRTSPEC

Examples

- Exact client identifier:

```
1)  LUMAP   LU1     USERID,USER1
2)  LUMAP   LU2     NAME1.HOST1.COM
3)  LUMAP   LU3     1.2.3.4
```

  Client Identifier type USERID is required. If not specified, USER1 is assumed to be a link or interface name.

- Exact client identifier in a group definition:

```
LUGROUP   LUGRP1    LU100..LU199   ENDLUGROUP
LUGROUP   LUGRP2    LU200..LU299   ENDLUGROUP
LUGROUP   LUGRP3    LU300..LU399   ENDLUGROUP
```

```
    USERGROUP USRGRP1
      USER1 USER2 USER3
    ENDUSERGROUP
```

```
    HNGROUP  HNGRP1
         NAME2.HOST1.COM  NAME2.HOST3.COM
    ENDHNGROUP
```

```
    IPGROUP  IPGRP1
      1.2.3.5     1.2.3.6
      1.3.4.7..1.3.4.E
    ENDIPGROUP
```

```
4)  LUMAP   LUGRP1   USRGRP1
5)  LUMAP   LUGRP2   HNGRP1
6)  LUMAP   LUGRP3   IPGRP1
```

- Wild card match for client identifier in a group definition:

```
    USERGROUP USRGRP2
      USER%%  TCPU*
    ENDUSERGROUP
```

```
    HNGROUP  HNGRP2
         *.HOST2.COM  **.HOST3.COM
    ENDHNGROUP
```

```
    IPGROUP  IPGRP2
      255.255.0.0:2.3.0.0
      2001:0DB8:3:274C::0/80
    ENDIPGROUP
```

```
7)  LUMAP   LUGRP1   USRGRP2
8)  LUMAP   LUGRP2   HNGRP2
9)  LUMAP   LUGRP3   IPGRP2
```

- Exact destination:

```
10) DEFAULTAPPL    TSO        DESTIP,1.2.3.4
11) USSTCP         USSTAB1    LINK1
```

  Client Identifier type DESTIP is required. If not specified, destination IP address 1.2.3.4 is assumed to be a client IP address.

- Exact destination in a group definition:

```
    DESTIPGROUP DSTIPGRP1
      1.2.3.5  1.2.3.6
      79DA:10:3.4.9.5
      613D:10::9241..613D:10::C510
    ENDDESTIPGROUP
```

```
    LINKGROUP LINKGRP1
      LINK1 LINK2 LINK3
    ENDLINKGROUP
```

```
12) LUMAP LUGRP1 DSTIPGRP1
13) LUMAP LUGRP2 LNKGRP1
```

- Wild card match for destination in a group definition:

```
    DESTIPGROUP DSTIPGRP2
      255.255.0.0:1.4.0.0
    ENDDESTIPGROUP
```

```
    LINKGROUP LINKGRP2
      LINK* %LINK
    ENDLINKGROUP
```

```
14) LUMAP LUGRP1 DSTIPGRP2
15) LUMAP LUGRP2 LNKGRP2
```

- Null client ID

```
16) DEFAULTAPPL     TSO
    LINEMODEAPPL    CICS
```

```
    USSTCP          USSTAB1
    INTERPTCP       INTTAB1
    DEFAULTLUS
      LU01..LU99
    ENDDEFAULTLUS
```

NULL is a single Client Identifier. The order of the examples has no significance. If DEFAULTAPPL and USSTCP mapping statements both have the NULL Client Identifier, the DEFAULTAPPL will be used regardless of order. For more information, see "Resolving DEFAULTAPPL and USS table conflicts" on page 622.

Object assignment examples

A client can be known by several different Client Identifiers. These Client Identifiers are used to assign as many Objects as possible to the connection based on the profile mapping statements. Telnet starts with the highest priority Client Identifier of the client and assigns all Objects mapped by that Client Identifier. If all 11 Objects are not assigned, Telnet uses the next highest priority Client Identifier (for prioritization details, see "Client Identifier selection rules" on page 612) and assigns all Objects mapped by that Client Identifier. This Object assignment process continues by using lower and lower priority Client Identifiers until all 11 Object types are found or until all of the matching Client Identifier mappings have been checked. If an Object is mapped by multiple Client Identifiers, only the Object mapped by the highest Client Identifier is used. It is unlikely all Objects are assigned to connections because not all Objects are always mapped. For example, many profiles do not contain PRTDEFAULTAPPL or INTERPTCP mapping statements. In this case, the printer default appl and Interpret table Objects will not be assigned.

Figure 89 on page 615 is a graphical representation of the following Telnet mapping statements. The numbered mapping statements correspond to the numbered buttons in the figure. The mappings that specify USERGROUP USGRP1 generate buttons 4 through 8 for exact user ID in a group and buttons 12 through 16 for wildcard user ID in a group.

```
LUGROUP      LUGRP1   LU01..LU10..FFNN    ENDLUGROUP
LUGROUP      LUGRP2   LU11..LU99..FFNN    ENDLUGROUP
PRTGROUP     PRTGRP1  PRT01..PRT10..FFFNN ENDPRTGROUP
PARMSGROUP   PGDBG    DEBUG DETAIL        ENDPARMSGROUP
PARMSGROUP   PGSCAN   SCANINTERVAL 10     ENDPARMSGROUP
PARMSGROUP   PGMTKO   TKOSPECLU  7        ENDPARMSGROUP
PARMSGROUP   PGALL    DEBUG DETAIL
                      SCANINTERVAL 10
                      TKOSPECLU  7        ENDPARMSGROUP
MONITORGROUP MONGRP1  NODYNAMICDR         ENDMONITORGROUP
USERGROUP    USGRP1   PAYUSR1   PAYUSR*   ENDUSERGROUP
HNGROUP      HNGRP1   USER1.GROUP3.COM
                      USER5.GROUP3.COM    ENDHNGROUP

   (1) PARMSMAP       PGALL    USERID,PAYUSR1
   (2) LINEMODEAPPL   TSO      9.9.9.9
   (3) PARMSMAP       PGDBG    9.9.9.9

(4,12) DEFAULTAPPL    PAYROLL  USGRP1
(5,13) PRTDEFAULTAPPL PAYPRT   USGRP1
(6,14) LUMAP          LUGRP1   USGRP1   SPECIFIC
(7,15) PRTMAP         PRTPGRP1 USGRP1   SPECIFIC
(8,16) PARMSMAP       PGTKO    USGRP1

   (9) USSTCP         USSTABHN HNGRP1
  (10) LUMAP          LUGRP2   HNGRP1   GENERIC
  (11) PARMSMAP       PGSCAN   HNGRP1

  (17) INTERPTCP      INTTAB1  LINK1
  (18) MONITORMAP     MONGRP1  LINK1
  (19) DEFAULTAPPL    TPX1
  (20) USSTCP         USSTAB1
```

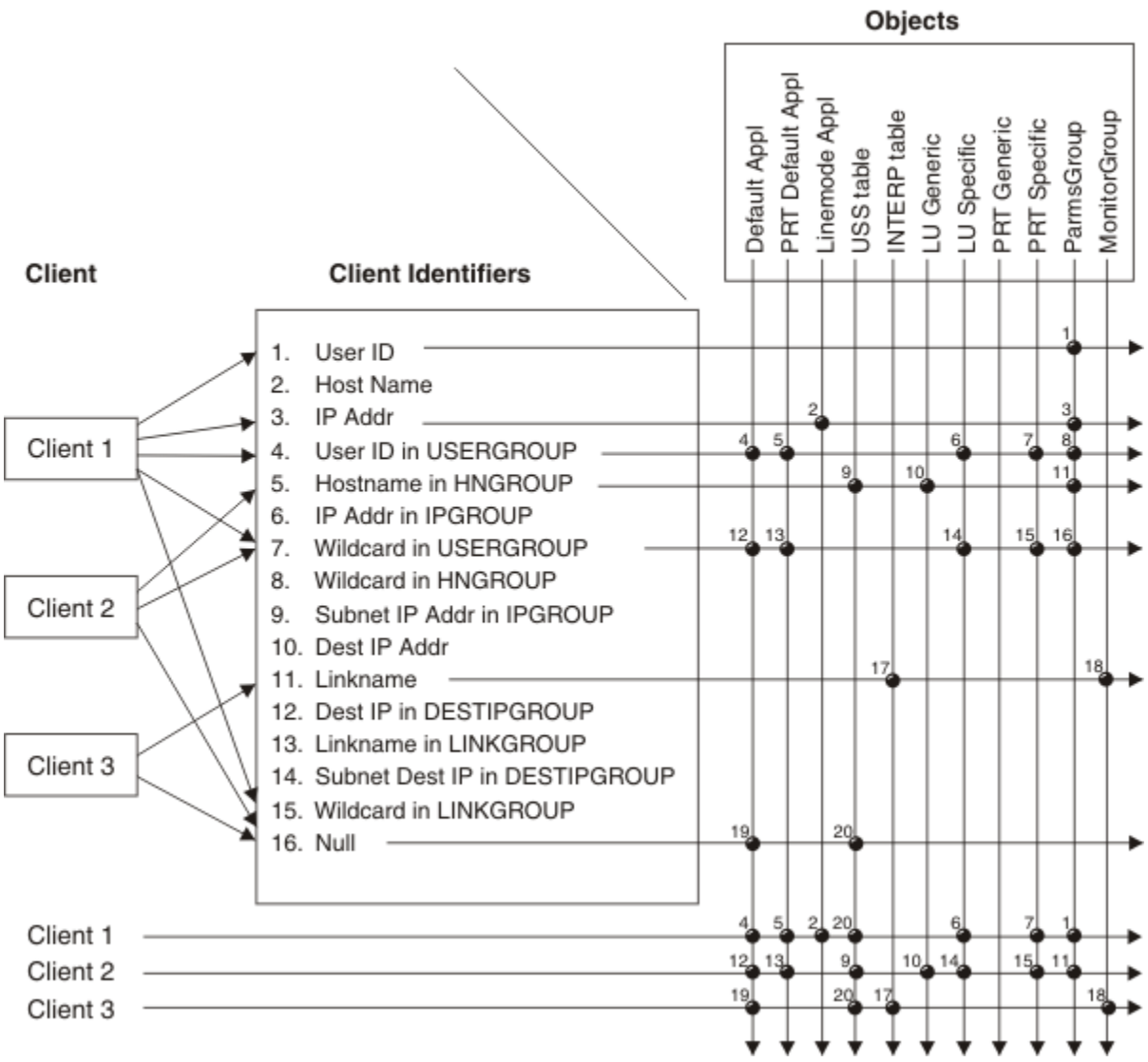


Figure 89. Search method

Client mappings

This example uses the following assumptions:

- Client 1 connects from IP address 9.9.9.9 using client authentication and is assigned PAYUSR1. The client does not have a host name ending in GROUP3.COM and does not have a link name LINK1.
- Client 2 connects from IP address 9.1.1.1 using client authentication and is assigned PAYUSR5. The client has the host name USER5.GROUP3.COM and does not have a link name LINK1.
- Client 3 connects from IP address 9.2.2.2 without client authentication and has the host name USER3.GROUP1.COM. The client connects to link name LINK1.

Based on Figure 89 on page 615, the clients are assigned objects as shown in Table 28 on page 616.

Table 28. Client mappings

Button	Object type	Name	Mapping results by client
1	ParmsGroup	PGALL	1: Assigned, exact user ID match 2: Ignored, no exact user ID match 3: Ignored, no exact user ID match
2	Linemode Appl	TSO	1: Assigned, exact IP address match 2: Ignored, no exact IP address match 3: Ignored, no exact IP address match
3	ParmsGroup	PGDBG	1: Already assigned by button 1 2: Ignored, no exact IP address match 3: Ignored, no exact IP address match
4	Default Appl	PAYROLL	1: Assigned, exact user ID in group match 2: Ignored, no exact user ID match 3: Ignored, no exact user ID match
5	PRT Default Appl	PAYPRT	1: Assigned, exact user ID in group match 2: Ignored, no exact user ID match 3: Ignored, no exact user ID match
6	LU Specific	LUGRP1	1: Assigned, exact user ID in group match 2: Ignored, no exact user ID match 3: Ignored, no exact user ID match
7	PRT Specific	PRTGRP1	1: Assigned, exact user ID in group match 2: Ignored, no exact user ID match 3: Ignored, no exact user ID match
8	ParmsGroup	PGTKO	1: Already assigned by button 1 2: Ignored, no exact user ID match 3: Ignored, no exact user ID match
9	USS table	USSTABHN	1: Ignored, no exact host name match 2: Assigned, exact host name in group match 3: Ignored, no exact host name match
10	LU Generic	LUGRP2	1: Ignored, no exact host name match 2: Assigned, exact host name in group match 3: Ignored, no exact host name match
11	ParmsGroup	PGSCAN	1: Ignored, no exact host name match 2: Assigned, exact host name in group match 3: Ignored, no exact host name match
12	Default Appl	PAYROLL	1: Already assigned by button 4 2: Assigned, wildcard user ID match 3: Ignored, no wildcard user ID match

Table 28. Client mappings (continued)			
Button	**Object type**	**Name**	**Mapping results by client**
13	PRT Default Appl	PAYPRT	1: Already assigned by button 5 2: Assigned, wildcard user ID match 3: Ignored, no wildcard user ID match
14	LU Specific	LUGRP1	1: Already assigned by button 6 2: Assigned, wildcard user ID match 3: Ignored, no wildcard user ID match
15	PRT Specific	PRTGRP1	1: Already assigned by button 7 2: Assigned, wildcard user ID match 3: Ignored, no wildcard user ID match
16	ParmsGroup	PGTKO	1: Already assigned by button 8 2: Ignored, no wildcard user ID match 3: Ignored, no wildcard user ID match
17	INTERP table	INTTAB1	1: Ignored, no link name match 2: Ignored, no link name match 3: Assigned, link name match
18	MonitorGroup	MONGRP1	1: Ignored, no link name match 2: Ignored, no link name match 3: Assigned, link name match
19	Default Appl	TPX1	1: Already assigned by button 4 2: Already assigned by button 9 3: Assigned, NULL Client ID match
20	USS table	USSTAB1	1: Assigned, NULL Client ID match 2: Already assigned by button 9 3: Assigned, NULL Client ID match

LU name mapping statements

Every connection must be represented by an LU name before a session can be initiated. The time of LU assignment depends on the connection type. In general, for TN3270E clients, the LU name is assigned early during connection negotiation before an application name is known. For all other types of clients, the LU name is assigned immediately after application name selection. For details and exceptions to this rule, see “Advanced LU name mapping topics” on page 624. Mapping statements define which LU name is assigned to the connection.

DEFAULTLUS

The simplest way to assign LUs is to create a default LU group that Telnet can use for all terminal connections. DEFAULTLUS is a combination statement that defines the LUs in a default group and maps the group to the NULL Client Identifier. If the client's Client Identifiers do not match any LU mapping statements, the client is identified by the NULL Client Identifier and will be assigned LUs from the default group.

For example, use the following statement to create an LU group with a numeric range of LUG1001 to LUG1100. When Telnet assigns an LU to a terminal connection, it will assign the next available LU from that group of 100 LUs.

```
DEFAULTLUS  LUG1001..LUG1100..FFFFNNN  ENDDEFAULTLUS
```

By default, Telnet uses a sequential selection method to assign LUs from the LU group. No LU name will be reused until all the names in the group have been used. Specifying NOSEQUENTIALLU changes the selection process to always start at the beginning and find the first name available. If the range is large and a large number of LUs are already assigned, NOSEQUENTIALLU might degrade LU lookup performance.

DEFAULTPRT

The DEFAULTPRT statement is used to create a default LU pool that Telnet will use for all printer connections. For example, use the following statement to create an LU group with a numeric range of PRTG1001 to PRTG1100. When Telnet assigns an LU to a printer connection, it will assign the next available LU from that group.

```
DEFAULTPRT  PRTG1001..PRTG1100..FFFFFNNN  ENDDEFAULTPRT
```

LUMAP, PRTMAP, LUGROUP, PRTGROUP

The LUMAP and PRTMAP statements allow you to map LUs to connections based on the Client Identifier for terminal emulators and printer emulators, respectively. For example, use the following statements to map LU name LUT001 to any terminal client identified by the client IP address 1.1.1.1 and map LU name PRT001 to any printer client identified by client IP address 2.2.2.2.

```
LUMAP  LUT001  1.1.1.1
PRTMAP PRT001  2.2.2.2
```

A local or shared LU group can be used when it is not necessary to have an exact LU name to Client Identifier match. For example, use the following statements to create a terminal LU group and a printer LU group, and map both groups to the Client Identifier IPGPAY. When a terminal client connects, Telnet will assign an LU from LUGRP1. When a printer client connects, Telnet will assign an LU from PRTGRP1.

```
LUGROUP   LUGRP1    LUT101..LUT400..FFFNNN   ENDLUGROUP
PRTGROUP  PRTGRP1   PRT101..PRT400..FFFNNN   ENDPRTGROUP

IPGROUP IPGPAY   255.255.0.0:9.8.0.0  ENDIPGROUP

LUMAP    LUGRP1   IPGPAY
PRTMAP   PRTGRP1  IPGPAY
```

If these same LUs can be mapped by more than one Telnet, put them into shared LU groups instead by adding an S to the object type as follows:

```
SLUGROUP   LUGRP1    LUT101..LUT400..FFFNNN   ENDSLUGROUP
SPRTGROUP  PRTGRP1   PRT101..PRT400..FFFNNN   ENDSPRTGROUP
```

After all 300 LUs are assigned, the next client connection request will fail. In this way, the LUGROUP Object can limit the number of clients connected at one time.

If a client connection is known by a Client Identifier that has an LU group mapping, only that mapping will be used to assign an LU name. The DEFAULTLUS group will not be used. It is used only in the case when no other LU mapping exists.

LU range specification

Telnet LU range rules allow for almost any type of LU range needed. Ranges can be alphabetic (A), numeric (N), alphanumeric (B), hexadecimal (X), or a complete wildcard (?), which includes alphanumeric and the three national characters (@,#,$). The range type can be different for each character position. Within the LU range, any character position can be fixed (F). To conform with VTAM LU naming convention,

the first character must be alphabetic or a national character. If the first character is a range, only the alphabetic range can be used.

An LU range is created by specifying a starting LU name, an ending LU name, and the range rules to be used. For example, the following statement creates a range from TCPM1000 to TCPM1100.

```
TCPM1000..TCPM1100..FFFFFNNN
```

The three components are:

- Starting LU name (TCPM1000)
- Ending LU name (TCPM1100)
- Range rules (FFFFFNNN)

All three components must be the same length, the Starting LU name overall must be lower than the Ending LU name, and each character position value must be appropriate for the specified range rule.

Tip: In the above example, the character 1 following the character M is defined as fixed because it cannot change. The range rule cannot specify N even though it seems to be part of the number range.

The ascending order of characters is 0-9, A-Z, @, #, $.

If the range rule is omitted, Telnet assumes the following style, where LowerRange and UpperRange must be all numeric or all alphabetic:

```
LuBase+LowerRange..LuBase+UpperRange
```

Numeric values are lower than alphabetic values to facilitate the use of hexadecimal ranges. The range rules are:

```
Range         Rule       Characters
--------------------------------------
Numeric       N          0-9
Alphabetic    A          A-Z
AlphaNumeric  B          0-9,A-Z
Hexadecimal   X          0-9,A-F
Wildcard      ?          0-9,A-Z,@,#,$
```

The maximum number of LUs per range is 4294967295 and the maximum number of LUs per group is 4294967295.

The creation of LU name values from the range specification begins at the Starting LU and increments the rightmost variable position first, moving to the left as each variable position reaches its range maximum. The process is like an odometer, except that each position can have different basing instead of all positions being base 10. For example, the following statement has 223 LU name entries.

```
LU555..LU777..FFNNN
```

The breakdown of the range is:

```
LU555->LU559,                                                 5
LU560->LU569, LU570->LU579, LU580->LU589, LU590->LU599,      40
LU600->LU699,                                               100
LU700->LU769, LU770->LU777                                   78
                                                            ===
Total  ---------------------------------------------------> 223
```

The LU names increment just as the numbers on an odometer would. A less intuitive case involves an alphabetic range of 1407 LU name entries.

```
LUCCC..LUEEE..FFAAA
```

The breakdown of the range is:

```
LUCCC->LUCCZ,                                                24
LUCDA->LUCDZ, LUCEA->LUCEZ, LUCFA->LUCFZ, ... LUCZA->LUCZZ,  598
```

```
LUDAA->LUDZZ,                                                  676
LUEAA->LUEDZ, LUEEA->LUEEE                                     109
                                                              ====
Total  ------------------------------------------------------> 1407
```

It is important to realize that these ranges do not break down in the following patterns:

```
LUCCC->LUCCE, LUCDC->LUCDE, LUCEC->LUCEE, ...
LU555->LU557, LU565->LU567, LU575->LU577, ...
```

It is an incorrect assumption that the LU name after LUCCE would be LUCDC. The correct LU name after LUCCE is LUCCF. The LU names increment to LUCCZ and the next name is LUCDA. When the rightmost position reaches the range maximum, the position to its left is incremented by one, and the rightmost position starts at the range beginning, not the character specified in the Starting LU name.

All range types are handled the same way. The position is incremented to its maximum value and then wraps to the beginning range value, not the specified Starting LU name value. By the same logic, the position is incremented to the ending range value and not the Ending LU name value.

All LU names increment the same way. A more complicated example mixes fixed and variable character positions with several different range types. The LU range has 39744 LUs.

```
LUAD1800..LUGD98FZ..FFAFNFXB
```

Calculating the number of LUs is easier if the fixed positions are removed. For purposes of calculating the number of LUs, the range is specified as follows:

```
A100..G9FZ..ANXB
```

This breaks down as follows:

```
A100->A10Z, D110->D11Z, ... A190->A19Z, A1A0->A1AZ ... A1F0->A1FZ    576
A200->A2FZ, A300->A3FZ, ... A900->A9FZ                              4608
B000->B9FZ, C000->C9FZ, ... F000->F9FZ                             28800
G000->G9FZ                                                          5760
                                                                   =====
Total ------------------------------------------------------------> 39744
```

SEQUENTIALLU

Telnet, by default, uses a sequential method to choose LUs from a group.

```
LUGROUP  LUGRP1
     LU001..LU120..FFNNN
     LU201..LU250..FFNNN
     LU240..LU280..FFNNN
     LU010..LU050..FFFNN
ENDLUGROUP
```

From the previous example, the first LU assigned is LU001, second is LU002, and so on. If five clients repeatedly connect and disconnect, they will be assigned new LUs farther into the range each time:

- When the end of the first range is reached, selection goes to the beginning of the second range.
- At the end of the second range, selection goes to the beginning of the third range.
- At the end of the third range, selection goes to the beginning of the fourth range.
- At the end of the fourth range, selection goes to the beginning of the first range again.

Telnet does not enforce an overall ascension in LU name selection. The selection process begins at the first name of the first range and progresses to the last name of the last range. In the example, after LU250 is assigned from range 2, LU240 from range 3 is attempted next. After LU280, LU010 is attempted. After LU050, the process starts over and LU001 is attempted.

The SEQUENTIALLU function can be turned off by coding NOSEQUENTIALLU. In this case, the five LUs that are repeatedly connecting and disconnecting would never use any LU names other than LU001, LU002, LU003, LU004, and LU005. NOSEQUENTIALLU might degradate LU lookup performance when a

large range is specified and only LUs at the end of the range are available. Every connection has to relearn that most of the LUs are already in use. SEQUENTIALLU allows Telnet to start its search near the last chosen LU where LUs are more likely to be available. SEQUENTIALLU and NOSEQUENTIALLU parameters can be coded at all three parameter block levels for different levels of granularity.

If several clients are connecting at the same time, the order of LU assignment might not be in exactly the same order as the connection IDs due to process timing between connection ID assignment and LU name assignment.

If single LU names are in a group with LU ranges, the single LU names are selected before any LU range names are selected, regardless of their order. In the example below, LUAAA, LUBBB, LUCCC, and LUDDD are all processed before any of the range LU names.

```
 Profile LUGROUP                              LUGROUP as used by Telnet


LUGROUP  LUGRP2                                 LUGROUP  LUGRP2
     LUAAA                                           LUAAA
     LU001..LU120..FFNNN                             LUDDD
     LU201..LU250..FFFNN                             LUBBB
     LUDDD                                           LUCCC
     LUBBB                                           LU001..LU120..FFNNN
     LU240..LU280..FFFNN                             LU201..LU250..FFFNN
     LU010..LU050..FFFNN                             LU240..LU280..FFFNN
     LUCCC                                           LU010..LU050..FFFNN
ENDLUGROUP                                      ENDLUGROUP
```

Application mapping statements

When a client connects, Telnet either immediately initiates a session request to an MVS host VTAM application or solicits the user for an application name.

DEFAULTAPPL

The DEFAULTAPPL mapping statement is used to assign an application name to the connection and immediately initiate a session with that application, and not solicit the user for an application name. The DEFAULTAPPL statement applies only to terminal emulators connecting in TN3270, TN3270E, or DBCSTRANSFORM mode. For example, use the following statement to map the default application PAYROLL to any TN3270 or TN3270E terminal client identified by the IPGROUP IPGPAY. When a TN3270 or TN3270E client connects, Telnet will immediately initiate a session to the PAYROLL application.

```
DEFAULTAPPL PAYROLL IPGPAY
```

PRTDEFAULTAPPL and LINEMODEAPPL

The PRTDEFAULTAPPL mapping statement is used to assign an application to a printer emulator client connecting in TN3270E mode. The LINEMODEAPPL mapping statement is used to assign an application to a client connecting in standard or binary LINE mode. For example, use the following statements to map the default application PAYPRINT to any TN3270E printer client identified by the IPGROUP IPGPAY and to map the default application TSO to any linemode client identified by the linkname LINK1. When the printer client connects, Telnet will immediately initiate a session to the PAYPRINT application. When a linemode client connects, Telnet will immediately initiate a session to the TSO application.

```
PRTDEFAULTAPPL PAYPRINT IPGPAY
LINEMODEAPPL  TSO  LINK1
```

The DEFAULTAPPL, PRTDEFAULTAPPL, and LINEMODEAPPL statements imply a basic ALLOWAPPL statement for the application name if no ALLOWAPPL or RESTRICTAPPL is explicitly coded.

USSTCP

If the user needs the ability to choose an application, you can create custom solicitor screens using unformatted system services (USS) message tables. These tables are mapped to clients using the USSTCP mapping statement. For example, use the following statement to map a USS table, USSTAB1, to any

TN3270 or TN3270E client identified by any link name that starts with LINK. When a TN3270 or TN3270E client connects, Telnet will immediately send a custom logon screen (USSMSG10) from the USS table.

```
LINKGROUP  LNKGRP1  LINK*   ENDLINKGROUP
USSTCP USSTAB1 LINKGRP1
```

Assembled USS tables used by VTAM can also be used by Telnet.

INTERPTCP

In some cases, the application name must be generated based on the name provided by the user or the name might be dependent on the LU name representing the client. The INTERPRET table can provide this function. Telnet uses the input from the USSMSG10 screen as input to the INTERPRET table translation list or uses the USSMSG10 input and the LU name as input to one of the INTERPRET table user-written exits. Because USS logon data is required input to the INTERPRET process, any client with an INTERPRET table mapping must also have a USS table mapping. For example, use the following statement to map an INTERPRET table, INTTAB1, to any TN3270 or TN3270E client identified by the linkname LINK1. When a TN3270 or TN3270E client connects to LINK1, Telnet will immediately send a custom logon screen (USSMSG10) from the USS table. The user responds with a USS logon command. LINK1 client input is then processed through the INTTAB1 INTERPRET table to derive an application name. Telnet uses the derived name to initiate a session.

```
LINKGROUP LNKGRP1
     LINK*
ENDLINKGROUP
USSTCP    USSTAB1 LNKGRP1
INTERPTCP INTTAB1 LINK1
```

Assembled interpret tables used by VTAM can also be used by Telnet.

If neither a default application nor a USS table is mapped to the connection, the Telnet solicitor screen is sent to the user. For a detailed discussion of the Telnet solicitor, USS table, and INTERPRET table, see “Using the Telnet solicitor or USS logon screen” on page 644.

Resolving DEFAULTAPPL and USS table conflicts

If both a default application and a USS table are mapped to the same Client Identifier, Telnet will use the default application to immediately initiate a session. If each is mapped by a different Client Identifier, the Object mapped by the higher priority Client Identifier is used. In all cases, any error messages are sent using the USS table messages. For example, if CICS and USSTAB1 are both mapped to destination IP address 1.1.1.1, Telnet will initiate a session with CICS and use the USS messages for any session setup errors.

If CICS is mapped to USERID USER1 and USSTAB1 is mapped to client IP address 5.5.5.5, Telnet will initiate a session with CICS and use the USS messages for any session setup errors.

If CICS is mapped to linkname LINK1 and USSTAB1 is mapped to hostname TEST1.IBM.COM, Telnet sends a USSMSG10 logon screen to the user. The USS messages will be used for any session setup errors. The default application mapping of CICS will never be used.

ALLOWAPPL

Telnet will not initiate a session for a solicited application name unless the name is allowed. The ALLOWAPPL statement is used to configure Telnet to allow the initiation request. For example, CICS01 and CICS02 are allowable names.

```
ALLOWAPPL CICS01
ALLOWAPPL CICS02
```

The ALLOWAPPL name can have a wildcard value by using an asterisk (*). For example, if there are no other CICS regions, these lines can be reduced to the following line:

```
ALLOWAPPL CICS*
```

All application names can be allowed by coding the following line:

```
ALLOWAPPL *
```

Default application names do not need to be explicitly allowed. However, if the default application issues a CLSDST-PASS to another application name for the session, the second application must be in the ALLOWAPPL list. For example, TSO is the default application for the NULL Client Identifier. TSO typically passes the session to TSO00001, TSO00002, and so on. The following default application mapping will initiate a session with TSO, but when TSO issues a CLSDST-PASS the new bind to Telnet will have TSO00001 as the application name.

```
DEFAULTAPPL TSO
```

Telnet will fail this session request because TSO00001 is not allowed. Add an ALLOWAPPL statement to allow the TSO* names as follows:

```
DEFAULTAPPL TSO
ALLOWAPPL   TSO*
```

RESTRICTAPPL

In addition to the ALLOWAPPL statement, Telnet provides more restrictive access to applications. The RESTRICTAPPL statement requires the user to enter a valid user ID and password before the application name is used to initiate a session. Specify the PASSWORDPHRASE parameter statement in Telnet configuration to expand the password length limit on the solicitor screen. You can use either a password or a password phrase with this option.

The user ID specified can be any valid user ID and does not need to be related to the user ID specified for the application. If you code the CERTAUTH option on the RESTRICTAPPL statement, the user does not need to supply a user ID if a client certificate is received and a user ID is derived from the client certificate. In this case, if the user ID derived from the client certificate matches a user ID on the RESTRICTAPPL statement, Telnet immediately initiates a session and does not request a password or password phrase from the user.

For example, use the following statement to allow users USER1, USER2, USER3, USER4, and USER5 access to the PAYROLL application. At the solicitor screen, the user enters USER1, their password or password phrase, and the PAYROLL application name. Telnet verifies that USER1 and the password or password phrase are valid and then immediately initiates a session with PAYROLL.

```
RESTRICTAPPL PAYROLL
   USER USER1
   USER USER2
   USER USER3
   USER USER4
   USER USER5
```

Like ALLOWAPPL, the application name can have a wildcard value by using an asterisk (*). The USER value can also have a wildcard value by using an asterisk. The user ID and password or password phrase combination is used by Telnet to verify the password or password phrase given for that user ID. In no way is the user ID and password or password phrase used by the application. No matter how the application name request arrived at the server (from DEFAULTAPPL or USSMSG10), Telnet uses the solicitor screen to prompt for the user ID and password or password phrase. After the user ID is validated and a password or password phrase is obtained, Telnet submits the user ID and password or password phrase pair for authorization to a security program such as RACF. The user ID and password or password phrase check authorizes the client to connect to the application through Telnet. The application itself might also ask for a user ID and password or password phrase pair that can be completely different than the pair entered at the Telnet solicitor screen. The user ID and password or password phrase pair that is entered at the Telnet solicitor screen is not passed to the host application. The user ID and password or password phrase pair is solicited only after the user enters an application name on the solicitor (or USSMSG10) screen. If a second application is reached through the original application using CLSDST-PASS, the second

application is verified and Telnet will solicit a new user ID and password or password phrase pair if necessary.

When searching for a match with the input application name, Telnet will find the most specific match whether it is on the ALLOWAPPL or RESTRICTAPPL statement. If each statement has the same name specified, the RESTRICTAPPL entry is used. For example, TSO has its own user ID and password or password phrase requirement and probably does not need the additional Telnet security check. However, the Telnet security check may be needed for all other applications. This example can be supported with the following statements.

```
RESTRICTAPPL   *
  USER   *
ALLOWAPPL   TSO*
```

Connection parameters mapping statement

Connection parameters are typically defined once at the port level. Sometimes it is useful to have different connection parameters depending on the Client Identifier. The PARMSGROUP and PARMSMAP statements allow connection parameters to be mapped at the Client Identifier level. This level of granularity applies to almost all parameters. See z/OS Communications Server: IP Configuration Reference for a list of Telnet parameters allowed in the PARMSGROUP block.

Assume the PAYROLL department is assigned the highest level of security and connections are being monitored with summary debug messages, general users are assigned negotiable security, and inventory employees are experiencing intermittent problems with Telnet connections that require detailed debug messages for resolution. The following statements assign the security and debug levels to the areas needed and do not affect other areas. See “Transport Layer Security” on page 598 for security information and “Telnet diagnostic tools” on page 580 for debug information.

```
HNGROUP   HNGINV
  **.GROUP3.COM
ENDHNGROUP
IPGROUP   IPGPAY
  255.255.0.0:2.2.0.0
ENDIPGROUP
IPGROUP   IPGGEN
  255.0.0.0:2.0.0.0
ENDIPGROUP
PARMSGROUP   PRMGDBG
  DEBUG DETAIL
ENDPARMSGROUP
PARMSGROUP   PRMGSEC1
  CONNTYPE SECURE
  DEBUG SUMMARY
ENDPARMSGROUP
PARMSGROUP   PRMGSEC2
  CONNTYPE NEGT
ENDPARMSGROUP
PARMSMAP   PRMGDBG  HNGINV
PARMSMAP   PRMGSEC1 IPGPAY
PARMSMAP   PRMGSEC2 IPGGEN
```

Advanced LU name mapping topics

Beyond the basic LU mapping statements, there are several functions available to the advanced user. This topic includes the following subtopics:

- Generic and Specific connection requests
- Mapping groups to Client Identifiers
- LU name assignment user exit
- Associated printer function
- Map default application and ParmsGroup by LU group
- Multiple LUMAP statements for one Client Identifier
- Keep LU for the Client Identifier

- LU group capacity warning
- LU mapping by application name
- LU mapping selection rules
- LU mapping with multilevel security active

Generic and Specific connection requests

There are three types of Telnet connection requests that dictate how Telnet chooses a name to represent the client. They are Generic requests, Specific requests, and associated printer requests. For details about associated printer requests, see “Associated printer function” on page 627. Most connection requests are Generic requests.

For Generic requests, Telnet has complete control over LU name assignment using the Generic mapping statements as a reference. All linemode and TN3270 connections use only Generic requests, and TN3270E terminal and printer emulators use Generic requests as their default request type. Specific mapping statements are ignored by Generic requests.

For Specific requests, the TN3270E client specifies the LU name to be used. Telnet validates the name using the Specific mapping statements as a reference. Requesting a Specific LU name allows a client to be assigned the same LU every time. This is important if the host application is LU name dependent, and the client does not have a constant Client Identifier to use for mapping an LU name. It is also important to block Telnet from assigning these LUs to Generic requests. That is why Specific mapping statements are ignored by Generic requests. If a Specific mapping does not find an LU match, Generic mapping statements are checked. Telnet confirms or denies the request during negotiation. If the LU mapping algorithms reject the client choice, Telnet sends a device type reject to the client. Most clients then notify the end user that the requested LU name is not valid or is already in use.

Default LU groups

DEFAULTLUS and DEFAULTPRT are default local LU groups for Generic requests from terminal and printer emulators. SDEFAULTLUS and SDEFAULTPRT are default shared LU groups for Generic requests from terminal and printer emulators. DEFAULTLUSSPEC and DEFAULTPRTSPEC are default local LU groups for Specific requests from terminal and printer emulators. SDEFAULTLUSSPEC and SDEFAULTPRTSPEC are default shared LU groups for Specific requests from terminal and printer emulators. DEFAULTLUS and DEFAULTPRT are explained in “LU name mapping statements” on page 617. Like the Generic pools, the Specific pools are checked only if there is no other LU mapping statement match. For example, use the following statements to create a terminal LU group with a numeric range of LUS1001 to LUS1100 and a printer LU group with a numeric range of PRTS1001 to PRTS1100. When Telnet receives a Specific connection request from a terminal, it will verify that the requested LU name is within the range specified.

```
DEFAULTLUSSPEC  LUS1001..LUS1100..FFFFNNN  ENDDEFAULTLUSSPEC
DEFAULTPRTSPEC  PRTS1001..PRTS1100..FFFFFNNN  ENDDEFAULTPRTSPEC
```

The sequential selection rules do not apply to Specific requests.

Mapping groups to Client Identifiers

The LUMAP and PRTMAP statements allow LUs to be mapped based on a Client Identifier. The LU group can be mapped Generically or Specifically. The default mapping is Generic. The keyword SPECIFIC must be coded to define a Specific mapping.

For example, use the following statements to create two LU groups. Map one group Generically to the IP group IPGPAY and map the other group Specifically to the same Client Identifier. When a Generic connection request is received, Telnet will assign the next available LU from LU group LUGRPGEN. When a Specific connection request is received, Telnet will verify the requested LU name is included in the LU group LUGRPSPC. If it is, Telnet will assign the LU name to the connection.

```
LUGROUP   LUGRPGEN  LUG101..LUG400..FFFXXX   ENDLUGROUP
LUGROUP   LUGRPSPC  LUS001..LUS100..FFFXXX   ENDLUGROUP

IPGROUP IPGPAY   255.255.0.0:9.8.0.0  ENDIPGROUP
```

```
LUMAP    LUGRPGEN   IPGPAY
LUMAP    LUGRPSPC   IPGPAY  SPECIFIC
```

Generic request connections can be assigned LUs only from Generically mapped LU groups. If no Generic mapping exists, the DEFAULTLUS group is checked. No Specific group is checked. This safeguards the Specific LU names from being used by Generic requests.

For Specific requests, Telnet first checks to see if the LU is in a Specifically mapped LU group. If the LU name is not found, the Generically mapped groups are searched. If neither LU group type contains the requested LU name, the connection request is rejected. The DEFAULTLUSSPEC group is not checked in this case because LU group mappings exist. If no LU group mappings exist, only the DEFAULTLUSSPEC group is checked. If the LU name is not found, the connection request is rejected. The Generic DEFAULTLUS group is not checked.

In addition to requesting an exact LU name, the client can request an LU group name. Telnet first searches within the mapped groups, assuming the name is an exact LU name. If that search fails, Telnet then checks the requested name against mapped Specific LU group names and then checks the name against mapped Generic LU group names. If the group name is found, the next available LU in the group is assigned using sequential LU selection unless it has been turned off.

The LU group itself is not defined as Generic or Specific. Rather, the LU group is *mapped* Generically or Specifically. It is possible to map the same LU group both Generically and Specifically. IBM recommends that you do not map the same group Generically and Specifically unless you are an advanced user.

LU name assignment user exit

Most LU assignment requirements can be satisfied using the Telnet LU group and LU mapping statements. However, there are cases when the LU assignment requirements are so specific that Telnet cannot satisfy them. In these cases, the LU name assignment user exit might be the solution. The LU name exit is defined like an LUGROUP and is mapped the same way LUGROUPs are mapped. The LUGROUP is defined as an exit by specifying *,EXIT* immediately after the LUGROUP name. For LUGROUPs, Telnet selects an LU from the group, verifies its availability, and assigns the LU to the connection. For LU name exits, Telnet calls the user-written assembler program passing a parameter list that contains client and other information. The program creates the LU name, places it in the parameter list, and returns control to Telnet. Telnet will then verify the LU name's availability and assign the LU to the connection. An LU name exit cannot be dynamically updated. After it is loaded, it remains unchanged until Telnet is recycled. To make a change without recycling Telnet, the exit name must be changed. The new name can then be added on a mapping statement. For a detailed description of the parameter list and coding requirements for the Telnet LU exit, see z/OS Communications Server: IP Configuration Reference.

Version 2 of the LU exit allows the exit to override Telnet profile assigned USS (3270 or SCS format) and interpret tables. The values assigned by the Telnet profile or blanks are passed into the exit. The exit can override any of the three values and Telnet will use the new values. For details on Telnet LU exit setup, see z/OS Communications Server: IP Configuration Reference.

In addition to client information, the parameter list includes any LU names or ranges that were coded in the LUGROUP, and the requested application name if known. Telnet does not use the LU list. The LUs specified can be used as seed values if the LU name exit wants to use them. The LUGROUP can be defined without any LUs specified. However, if the exit is used with multilevel security, at least one LU must be specified so the LUGROUP can be assigned a security label. For details, see “LU mapping with multilevel security active” on page 634.

The LU name exit is called when the LU is assigned, when the LU is released, when the LU is deactivated, and when the LU is activated. A different function code is used for each type of call. If you do not need a certain function, like tracking deactivated LUs, the LU name exit can be written to ignore the function code. Telnet allows only one connection at a time to use the LU name exit, which serializes its use in case any local tables are maintained in the exit.

As an example, assume LU names are to be assigned based on client port number and application requested. The SIMCLIENTLU parameter is used to postpone TN3270E LU assignment until the application name is known. The parameter list includes the client port number and the requested

application name. In this case, no seed LU names are needed. The LU name exit will create LU names based on the port number and application name in the parameter list.

```
LUGROUP  LUEXIT1,EXIT
ENDLUGROUP
```

In another example, assume that the clients specify LUGROUP names that match the default applications on the LUMAP statement. The LU names are to be created based on the last two numbers of the IP address and a prefix that identifies the application. For example, TSO, IMS, and CICS are three current applications, and the prefix for each is TS, IM, and CI, respectively. The connection from IP address 9.1.240.111 specifies LUGROUP LUGTSO and is assigned the name TS240111. The connection from IP address 9.1.240.212 specifies LUGROUP LUGIMS and is assigned the name IM240212. The connection from IP address 9.1.89.7 specifies LUGROUP LUGTSO and is assigned the name TS089007. Three LU name exits are required (LUGTSO, LUGIMS, LUGCICS), but they are all functionally equivalent. The LU name specified in the LUGROUP statement is passed to the LU name exit as part of the parameter list, and that name is used by the exit as the prefix. The client IP address is also in the parameter list. To force Telnet to call each exit, the LU name exit must return a nonzero return code when the LU name sent by the client does not match the LUGROUP name. The LU name exit combines the prefix with the last portions of the IP address to create an LU name. The following statements can be used to support this scenario.

```
IPGROUP  IPGRP1  0.0.0.0:0.0.0.0  ENDIPGROUP  ; Matches all connections


LUGROUP  LUGTSO,EXIT    TS       ENDLUGROUP
LUGROUP  LUGIMS,EXIT    IM       ENDLUGROUP
LUGROUP  LUGCICS,EXIT   CI       ENDLUGROUP

LUMAP  LUGTSO  IPGRP1  DEFAPPL  TSO
LUMAP  LUGIMS  IPGRP1  DEFAPPL  IMS
LUMAP  LUGCICS IPGRP1  DEFAPPL  CICS
```

Capacity checks cannot be performed because Telnet has no way of knowing how many total LUs are available in the LU name exit.

Associated printer function

The associated printer function allows a printer emulator to specify an active LU terminal name during connection negotiation. Telnet understands this special request and knows to assign a printer LU name that is associated with the requested terminal LU name. You establish the association by linking a pool of terminal LUs (LUGROUP or SLUGROUP) with a pool of printer LUs (PRTGROUP or SPRTGROUP). The groups are linked with the LUMAP statement. The printer LU group name is linked to the terminal LU group name by adding the PRTGROUP name on the LUMAP statement.

The two LU groups must have the same number of LUs defined so the LUs can be paired. The groups must have the same number of single LU names, the same number of LU ranges, and the same number of LU names in each range. If the groups do not have the same number of LUs defined, error messages will be produced during profile processing.

After the groups are linked, Telnet assigns the *n*th printer LU to a printer connection that requests association with the *n*th terminal LU. For example, a CICS table might specify that if terminal LU1 is requesting printer function, the output should be routed to printer PRT1. Within CICS, LU1 and PRT1 are associated with each other. Use the following statements to set up printer association.

```
LUGROUP  LUGCICS   LU1..LU9          ENDLUGROUP
PRTGROUP PRTCICS   PRT1..PRT9        ENDPRTGROUP
IPGROUP  IPGRP9    255.0.0.0:9.0.0.0  ENDIPGROUP
LUMAP  LUGCICS  IPGRP9  GENERIC  PRTCICS
```

To use shared groups, change the first two lines to indicate that the groups are shared:

```
SLUGROUP  LUGCICS   LU1..LU9          ENDSLUGROUP
SPRTGROUP PRTCICS   PRT1..PRT9        ENDSPRTGROUP
```

Rule: If you are using sysplex distributor to distribute connections across Telnet servers with associated printers, regardless of whether the LUs are in local groups or shared groups, you must use timed client affinity to ensure that the clients connect to the same Telnet server.

If the terminal connection matches a DEFAULTAPPL or LUMAP-DEFAPPL mapping statement, Telnet immediately initiates a session request for the specified application. If the printer connection matches a PRTDEFAULTAPPL mapping statement, Telnet immediately initiates a session request for the specified application. Building on the previous example, use the following statements to set CICS as the default application for both the terminal and printer connections:

```
LUMAP LUGCICS IPGRP9 GENERIC DEFAPPL CICS PRTCICS
PRTDEFAULTAPPL CICS
```

Neither the LU group nor the printer group can be an LU exit group. If either is an LU exit, the mapping statement will be rejected.

Drop the printer connection when dropping the terminal connection

In many cases, the associated printer connection should be dropped when the terminal connection is dropped. If you code the DROPASSOCPRINTER parameter, Telnet will monitor the terminal connection. When the terminal connection is dropped, Telnet will initiate the closing and dropping of the printer connection. The DROPASSOCPRINTER and NODROPASSOCPRINTER parameters can be coded at all three parameter block levels for different levels of granularity.

Map default application and ParmsGroup by LU group

The DEFAPPL option on the LUMAP statement allows a host VTAM application to be mapped with an LU name or LUGROUP name instead of using DEFAULTAPPL. The LUMAP-DEFAPPL combination is treated just like DEFAULTAPPL when a Client Identifier matches the LUMAP statement. The LUMAP-DEFAPPL combination also supports the LOGAPPL, FIRSTONLY, and DEFONLY parameters that are used by DEFAULTAPPL, PRTDEFAULTAPPL, and LINEMODEAPPL. The LUMAP-DEFAPPL combination is a powerful statement when used with multiple LUMAP statements for the same Client Identifier. If the LUMAP-DEFAPPL or PRTMAP-DEFAPPL statement is coded and the default application is not available, an error screen will be sent to the client whether or not MSG07 is coded.

The PMAP option on the LUMAP statement allows assignment of connection parameters based on LU or LU group name. When the LU is assigned, the parameter values specified in the PMAP PARMSGROUP will override the parameter value specified in TELNETGLOBALS, TELNETPARMS, or PARMSGROUP mapped to this connection's Client Identifier. For example, any client in subnet 9.0.0.0 that specifies the LU group LUGTSO will immediately have a session initiated to TSO with the LOGAPPL function, and the TIMEMARK time will be set for two hours instead of the default three hours.

```
IPGROUP    IPGRP9  255.0.0.0:9.0.0.0             ENDIPGROUP
LUGROUP    LUGTSO  TCPTSO01..TCPTSO99..FFFFFFNN  ENDLUGROUP
PARMSGROUP PGRPT2  TIMEMARK 7200                 ENDPARMSGROUP
LUMAP     LUGTSO  IPGRP9  DEFAPPL TSO LOGAPPL  PMAP PGRPT2
```

An LUMAP-DEFAPPL defined default application name is always used if specified, regardless of USSTCP mappings. LUMAP-DEFAPPL has a higher priority than any DEFAULTAPPL or USSTCP. The connection parameters assigned using LUMAP-PMAP will override any other setting of the parameters. However, not all parameters have a meaningful use by the time the LU is assigned. For example, NOTN3270E controls whether or not Telnet should negotiate for TN3270E. That negotiation is done before LU assignments. For information on which parameters can be properly applied with LUMAP-PMAP, see the parameter table in z/OS Communications Server: IP Configuration Reference.

The PRTMAP statement supports PRTMAP-DEFAPPL and PRTMAP-PMAP in the same manner as LUMAP-DEFAPPL and LUMAP-PMAP.

Multiple LUMAP statements

Another feature of Specific LU name requests is that the client can specify an LUGROUP name, and Telnet will assign an available LU from that pool. This capability is useful when different applications require

different LU naming schemes, but each user client does not need to use an exact LU name for each emulator. For example, an administrator can create three pools, one for each of three applications. Only three client emulators need to be set up. One for TSO which requests LU name LUTSO, one for CICS which requests LUCICS, and one for IMS which requests LUIMS. Assume the general users are in subnet 3.0.0.0. Any client connecting with a Client Identifier of IPGGEN can be set up to issue a Specific request for LU pool LUTSO, LUCICS, or LUIMS, and will be assigned an LU from the appropriate pool.

After an LU is assigned, the DEFAPPL option will cause Telnet to immediately issue a session request for the appropriate application. If LOGAPPL is coded and the application is not active, VTAM will continue session initiation after the application is active.

In most cases, DEFAPPL on multiple Generic LUMAPs is not useful. LUs are assigned in order starting with the first LUMAP statement. One case that may be useful is if an application has a user limit but can be cloned. Assume the INVENTRY application can support only 20 users but can be cloned. Multiple LUMAPs with DEFAPPL will direct the first 20 HNGINV clients to INVENTRY, the next 20 HNGINV clients to INVENTR2, and the next 20 HNGINV clients to INVENTR3.

```
IPGROUP  IPGGEN
     255.0.0.0:3.0.0.0
ENDIPGROUP
LUGROUP  LUTSO   TSO00001..TSO00999  ENDLUGROUP
LUGROUP  LUCICS  CICS0001..CICS0999  ENDLUGROUP
LUGROUP  LUIMS   IMS00001..IMS00999  ENDLUGROUP
HNGROUP  HNGINV
  *.INVDEPT.COM
ENDHNGROUP
LUGROUP  LUGINV1 LUINV01..LUINV20    ENDLUGROUP
LUGROUP  LUGINV2 LUINV21..LUINV40    ENDLUGROUP
LUGROUP  LUGINV3 LUINV41..LUINV60    ENDLUGROUP
LUMAP   LUTSO     IPGGEN   SPECIFIC  DEFAPPL  TSO      LOGAPPL  FIRSTONLY
LUMAP   LUCICS    IPGGEN   SPECIFIC  DEFAPPL  CICS     LOGAPPL
LUMAP   LUIMS     IPGGEN   SPECIFIC  DEFAPPL  IMS      LOGAPPL
LUMAP   LUGINV1   HNGINV             DEFAPPL  INVENTRY LOGAPPL  DEFONLY
LUMAP   LUGINV2   HNGINV             DEFAPPL  INVENTR2 LOGAPPL  DEFONLY
LUMAP   LUGINV3   HNGINV             DEFAPPL  INVENTR3 LOGAPPL  DEFONLY
```

Pool name specification is a powerful mapping method because multiple LUMAP statements with different Objects can be used for a single Client Identifier.

Keep LU for the Client Identifier

An LU name can be kept (or reserved) for a period of time so no other client is assigned that name. Only the same Client Identifier reconnecting to Telnet within the specified time can be assigned that LU name. After the specified time, the LU name is again available for any connection. This function is useful when the application does not clean up session information quickly and a released LU is quickly reassigned to another user by Telnet. The application thinks the new session is a continuation of the previous session but it is not. With KEEPLU, the LU will not be reassigned to a different Client Identifier for a period of time, long enough for the application to clean up its session information. The LU name is kept based on the highest Client Identifier by which the connection is known. It is either a User ID derived from a client certificate, a Hostname, or an IP address, respectively.

LU group capacity warning

An LU group capacity threshold can be specified on the LUGROUP, PRTGROUP, and default LU group statements. If specified, Telnet will check the number of LUs used in the group when an LU is assigned from the group. A message is issued when the group's in-use LU count is at or above the specified percentage of the total. After the message is issued, no other message is issued until the in-use count has dropped below the threshold by 10% of the total. For example, an LU group has 200 LUs with a capacity threshold of 80%. When the 160th LU is assigned, EZZ6007I is issued. Ten percent of the total in the group is 20. Therefore, after the number of in-use LUs has dropped to 140 or lower, another warning message will be issued when the in-use count rises to 160 again. If multiple LU groups have the same LU name, the only LU group checked is the group from which the LU is assigned to the client. The other LU

groups might go over their capacity limits, but notification will not be issued until an LU is taken from the group. Below are examples for setting the capacity warning.

```
LUGROUP         LUGRP1,80%    TCPLU000..TCPLUF9F..FFFFFXNX
PRTGROUP        PRTGRP1,60%   TCPRT000..TCPRTFFF..FFFFFXXX
DEFAULTLUS      ,75%          LU000000..LU999999..FFNNNNNN
DEFAULTPRTSPEC  ,90%          PRTDEFS1..PRTDEFS9
```

Capacity checking cannot be done for LU groups that are defined as LU name exits. During VARY TCPIP,*tnproc*,OBEYFILE command processing, all LU groups are checked for in-use LU counts and a capacity warning message is issued if needed.

Tip: Be sure to leave a blank space between the default LU group statement and the capacity (,*nnn*%). Do not leave a blank space between a group name and the capacity.

LU mapping by application name

In some cases, only certain LU names are eligible to be in session with the host application. Or only certain LU names are eligible to represent user IDs. The LU and LUG parameters on the ALLOWAPPL and RESTRICTAPPL statements provide this checking function and allow some LU name mapping based on application name. The LUG parameter can represent either an LUGROUP, a PRTGROUP, or a group that is a mixture of terminal and printer LUs so that both terminal and printer emulators can access the application. If single LUs are specified, they are assumed to be terminal LUs.

For example, assume the only LUs eligible to use the inventory set of applications are the LUs in the inventory LU pools. A new LUGROUP pool named LUGINVT contains LUs from LUGINV1, LUGINV2, and LUGINV3. The ALLOWAPPL statement requires that any session request to the inventory applications have an LU name defined in LUGINVT. The LUG parameter must be used carefully. When specified, Telnet must match the LU using both the common mapping algorithms and the mapping by application. For RESTRICTAPPL, assume security authorization is required to get to the PAYROLL application, and each of the PAYxx user IDs must map to a certain LU.

```
LUGROUP   LUGINV1  LUINV01..LUINV20  ENDLUGROUP
LUGROUP   LUGINV2  LUINV21..LUINV40  ENDLUGROUP
LUGROUP   LUGINV3  LUINV41..LUINV60  ENDLUGROUP
LUGROUP   LUGINVT  LUINV01..LUINV60  ENDLUGROUP
ALLOWAPPL   INVENTR*  LUG   LUGINVT
RESTRICTAPPL   PAYROLL
  USER  PAY01     LU    LUPAY01
  USER  PAY02     LU    LUPAY02
        (user pay03 through pay20 not listed)
```

The LU group specified on the LUG parameter cannot be an LU exit. If it is, the ALLOWAPPL statement is rejected. Multiple LUs can be assigned individually using the LU keyword or a single LU group can be assigned using the LUG parameter. LU and LUG cannot be mixed on a single statement and only one LUG entry per statement is permitted. LU assignment based on application is a convenient way to limit the access to applications. However, this increases mapping complexity significantly when LU mapping statements and connection types are part of the overall mapping equation. Non-TN3270E connections or TN3270E connections with NOTN3270E or SIMCLIENTLU specified do not keep the LU name assigned to the connection after a session is dropped. For these connection types, the user can establish a session with different application names even if different LU names are mapped to the application names with the ALLOWAPPL or RESTRICTAPPL-USER statement. However, LU mapping that is based on application name does not work well with TN3270E connections because the LU is assigned during connection negotiation before the correct application name is known. In all CLSDST-PASS cases, the LU name cannot change when switching from the first application to the second because the LU's ACB is not closed during the switch. If the LU mapping by application name requires an LU name switch, the new session attempt will be failed by Telnet.

TN3270 connections do not assign an LU to represent the client until an application name is chosen. Therefore, the LU and LUG parameters can be used as sole LU mapping statements for TN3270 connections. For example, assume no other mapping statements exist (LUMAP or DEFAULTLUS), and either no TN3270E connections will be used or SIMCLIENTLU has been specified. The following

ALLOWAPPL statements will map LUs to the appropriate application based on the application name chosen. The following RESTRICTAPPL statement will assign a single LU or LU pool to each user.

```
ALLOWAPPL   TSO*  LUG   LUGTSO
ALLOWAPPL   CICS  LUG   LUGCICS
ALLOWAPPL   IMS   LUG   LUGIMS
RESTRICTAPPL   APP*
  USER   USER1*   LUG  LUG10
  USER   USER01   LU   LU01
  USER   USER02   LU   LU02
```

Both of these assignment methods were very popular before TN3270E connections were introduced. TN3270E connections will likely achieve poor mapping results. An LU must be assigned during connection negotiation before the application name is known which will likely result in an LU mismatch later. TN3270E connections require that an LU mapping statement exist because an LU must be assigned to the connection during negotiations before an application name is known. Consider the following example:

```
DEFAULTLUS
  LU1 LU2 LU3 LU4
ENDDEFALTLUS
RESTRICTAPPL  APPL1
  USER  USER3  LU  LU3
ALLOWAPPL   APPL2  LU  LU4
```

Assume two TN3270 connections are started.

- Two solicitor screens appear.
- Specify APPL1, USER3, and a password. Telnet selects LU3 based on both the DEFAULTLUS and the RESTRICTAPPL statements.
- Specify APPL2. Telnet selects LU4 based on both the DEFAULTLUS and the ALLOWAPPL statements.

Assume two TN3270E connections are started.

- Two solicitor screens appear. Telnet assigns LU1 and LU2.
- Specify APPL1, USER3, and a password. Telnet fails the connection because of an LU mismatch.
- Specify APPL2. Telnet fails the connection because of an LU mismatch.

If LU name mapping by application name or user ID is wanted with TN3270E clients, the following three solutions are available:

- If the same application or user ID is always used at the same client, individual LUMAP statements can be used to map the correct LU name to each client. Then every connection request will result in the correct LU assignment for that client. The assumptions are that the client keeps the same Client Identifier and only one client exists per Client Identifier.
- Map the NOTN3270E parameter to clients to disable all TN3270E function in Telnet so those connections will be TN3270, not TN3270E. The drawback is that all TN3270E function is disabled. This includes printer function, Generic/Specific function, and SNA function to the client. The TN3270E and NOTN3270E parameters can be coded at all three parameter block levels for different levels of granularity.
- Mapping the SIMCLIENTLU parameter is a less severe solution. This function will send a dummy LU name of EZBSIMLU to all TN3270E clients issuing Generic connection requests to satisfy the negotiation but will not assign a Telnet LU until an application name is chosen. This alternative preserves printer function, Specific requests, and SNA function to the client. The drawback is the name sent to the client is not the name Telnet ultimately uses to represent the client. Printer association will not work for these TN3270E Generic connections and any emulator programming that depends on the LU name will be using the dummy LU name. The SIMCLIENTLU and NOSIMCLIENTLU parameters can be coded at all three parameter block levels for different levels of granularity.

LU mapping selection rules

LU mapping selection can become complicated because of the many variations of mapping statements, TN3270E versus TN3270 connections, Generic versus Specific connection requests, printer association,

and LU mappings based on application name. LU mapping is very different between TN3270E and TN3270 and will be discussed separately. But first, some general Mapping Rules for both TN3270E and TN3270 follow:

- If multiple LUMAP statements exist for a Client Identifier all Specific LUMAPs are searched (TN3270E only) and then all Generic LUMAPs are searched in the order they are listed in the profile.
- If the application is known during the LU lookup and the ALLOWAPPL or RESTRICTAPPL-USER statement has LUs listed, then the found LU must be in both the mapped LU group and in the application LU group.
- When an LU match is found, the search stops.
- Telnet performs database lookup for Objects based on the Client Identifier. TN3270E connections require an Early Lookup so Telnet can give the client an LU name during connection negotiation. In all cases a Complete Lookup is done when the application name is known. Telnet performs an Early Lookup and a Complete Lookup for TN3270E connections. Telnet performs only a Complete Lookup for TN3270 and Linemode connections.

TN3270E LU mapping

TN3270E connections require an Early Lookup during connection negotiation. Telnet will use as much information as is available to assign an LU to the client. However, the eventual application is not known at this time unless an LUMAP-DEFAPPL or DEFAULTAPPL statement defines the application name. After connection negotiations are complete, Telnet will either send a logon solicitor (or USSMSG10) screen to the client or will perform a Complete Lookup using the application name obtained from the LUMAP-DEFAPPL or DEFAULTAPPL statement. If Complete Lookup is successful, Telnet will begin session initiation. If a solicitor (or USSMSG10) screen is sent to the client, an application name must be entered, at which time Telnet will perform a Complete Lookup. If LU mapping is being done based on application name, a conflict might occur between the application LU mapping and the LU already assigned to the connection. For TN3270E, after an LU name is assigned during connection negotiation it can never change until the connection is dropped. The SIMCLIENTLU statement allows Telnet to assign LUs for TN3270E connections as though they were TN3270 connections. See “TN3270 LU mapping” on page 633 for mapping Generic TN3270E connection requests with SIMCLIENTLU. A request for a Specific LU from the Telnet Client will be treated as if SIMCLIENTLU were not specified. The exact lookup process for TN3270E (non-SIMCLIENTLU) is described below.

Early Lookup: An LU must be found during Early Lookup. LUMAP-DEFAPPL and DEFAULTAPPL statements are considered but not necessarily used. Possible lookup results are:

- An LU is found.
- An LU is not found, the connection is dropped.

Perform TN3270E Early Lookup in the following order. The process stops when LU lookup is successful. Printer connections use the same process, substituting PRTMAP and PRTDEFAULTAPPL.

- Check for LUMAP matches considering application lookup results and possible application-based LU mappings.
  1. For each Specific LUMAP used for Specific connection requests: If the Specific LUMAP has DEFAPPL, or DEFAULTAPPL was specified and the application lookup return code is either OK or USER_REQUIRED, then perform LU lookup.
  2. For each Generic LUMAP used for Specific or Generic connection requests: If the Generic LUMAP has DEFAPPL, or DEFAULTAPPL was specified and the application lookup return code is either OK or USER_REQUIRED, then perform LU lookup.
- Check for LUMAP matches without considering application lookup results.
  1. For each Specific LUMAP used for Specific connection requests: Ignore DEFAPPL and DEFAULTAPPL and perform LU lookup.
  2. For each Generic LUMAP used for Specific and Generic connection requests: Ignore DEFAPPL and DEFAULTAPPL and perform LU lookup.

- If LUMAP statements were not checked (different from checked but no match), use the appropriate Default LU pool considering application lookup results and possible application-based LU mappings. In this case the only relevant application is the DEFAULTAPPL, if specified. If the application lookup return code is either OK or USER_REQUIRED, then perform LU lookup.
- If LUMAP statements were not checked, try the appropriate Default LU pool without considering application lookup results. Perform LU lookup.

Complete Lookup: An application name is required for Complete Lookup. The application name is obtained from one of three sources in the order specified.

1. Input from the USER or VTAM (via CLSDST with OPTCD=PASS)
2. DEFAPPL parameter on the LUMAP statement
3. DEFAULTAPPL statement

Use the application name and the previously found LU to perform Complete Lookup. Possible lookup results are:

- The application is not valid.
- The application is valid (return code OK or USER_REQUIRED) for the existing LU.
- The application-based LU map does not match the already chosen LU.

TN3270 LU mapping

TN3270 connections perform Complete Lookup only after all information is known. LU lookup is not done during connection negotiation. Telnet will either send a solicitor (or USSMSG10) screen to the client or will perform Complete Lookup using the application name known through the LUMAP-DEFAPPL or DEFAULTAPPL statement. If Complete Lookup is successful, Telnet will begin session initiation. If not successful, the solicitor (or USSMSG07) screen is sent to the client without an LU being assigned to the connection or the connection is dropped. The LU is not assigned until the application name is valid. If the application name is a RESTRICTAPPL, the LU is not assigned until a user ID is specified. Application-based LU mappings have a very good chance of success due to the late LU mapping aspect of TN3270 connections. When SIMCLIENTLU is coded, Generic TN3270E connections have this same characteristic.

Complete Lookup: An application name is required for Complete Lookup. The application name is obtained from one of three sources in the order specified.

1. Input from the USER or VTAM (CLSDST with OPTCD=PASS)
2. DEFAPPL parameter on the LUMAP statement
3. DEFAULTAPPL statement

Use the application name to perform Complete Lookup. Possible lookup results are:

- The application is not valid.
- The application is valid but an LU is not found.
- The application is valid (return code OK) and an LU is found.
- The application LU map does not match the Client Identifier LU map.

If the application is not valid, no LU is assigned to the connection and an error message is sent to the client. If the application is valid, continue the LU lookup in the following order.

- Check for LUMAP matches considering application-based LU lookup results. Only Generic LUMAPs are searched. If the application lookup return code is OK, then perform LU lookup.
- If no LUMAP statements were used, check for application-based LU mappings. If the application lookup return code is OK and LUs are defined on the application statement, perform LU lookup.
- If no LUMAP or application-based LU mapping statements were used, use the DEFAULTLUS pool considering application lookup results. If the application lookup return code is OK, then perform LU lookup.

LU mapping with multilevel security active

Telnet can be in a multilevel secure environment that uses security labels. For more information about preparing for IP networking in a multilevel secure environment, see Chapter 4, "Preparing for IP networking in a multilevel secure environment," on page 191 and z/OS Planning for Multilevel Security and the Common Criteria. To ensure correct security label comparisons, Network Access Control (NAC) must also be active for Telnet. For more information about NAC, see "Network Access Control" on page 604.

If multilevel security is active, Telnet ensures the security label of the selected LU is compatible with the security label of the client.

- Telnet retrieves the security label of the client when the connection is accepted.
- Telnet assigns a security label to all LUGROUPs based on the first LU name in the group. The first single LU name in the group is used. If no single LU names exist, the first LU name within the first LU range is used.
  - If multilevel security is active, an LUGROUP EXIT is required to have at least one LU name in the group. The LU name is used to obtain a security label for the group. The name is passed to the exit in the parameter list and can be used or ignored by the exit.
  - A single LU name on a mapping statement is treated as an LUGROUP with one LU name. That LU name is used to obtain the security label for the LUGROUP created by Telnet.

When multilevel security is active, LU lookup uses the following process:

1. The security label of the client is compared with that of the mapped LUGROUP. If the group is compatible, Telnet searches for an available LU in the group. If not compatible, the LUGROUP is skipped.
2. Telnet retrieves the security label of the selected LU and compares it with the security label of the LUGROUP. If the selected LU is not compatible with the LUGROUP, the LU is deactivated and no other LU in the group is tried.
3. If the LUGROUP was not compatible or no LU was available, the steps are repeated for each mapped LUGROUP until an LU is found or all LUGROUPs are checked.

Advanced application topics

In addition to the basic function of facilitating session setup, Telnet supports several advanced functions such as:

- Connection information passed on the CINIT control vector 64 (CV64)
- Session initiation management (LOGAPPL, QINIT, FIRSTONLY, and DEFONLY)
- Check client connection and connection/session takeover
- Queueing sessions
- Disconnect on session error
- Bypass RESTRICTAPPL with CERTAUTH
- Allow printer sessions with RESTRICTAPPL
- Keeping the ACB open
- Express Logon Feature

Connection information passed on the CINIT control vector 64

During session establishment, a VTAM session initiation record called a CINIT is created and sent to the primary application. Attached to the CINIT is a control vector 64 (CV64) that is created by Telnet to provide additional information about the Telnet connection. The CV64 control vector contains four sub-vectors. The CV64 control vector and the four sub-vectors are all defined as key/length vectors.

- CV64 - Start of control vector

  - Key/Length (2 bytes)

    X'64*mm*', where *mm* is (*ii* + 2) + 4 + (*hh* +2 (if SBV85 present)) + 19 (if SBV86 present)
- SBV81 - Flags and IP address
  - Key/Length (2 bytes)

    X'81*ii*', where *ii* is 06 for IPv4, 18 for IPv6
  - IP address type (1 byte)

    04 - IPv4
    06 - IPv6
  - Flags (1 byte)

    80 - Reserved for future use
    40 - Reserved for future use
    20 - On if secure connection
    10 - On if secure connection flag is valid
    08 - On if takeover of the connection is possible
    04 - On if the TCP/IP stack supports IPv6 addresses and the client has an IPv4 address
    02 - On if the Telnet server is the z/OS Communications Server TN3270E Telnet server
    01 - On if the client is a TN3270E client and supports definite response
  - IP address (4 bytes if IPv4, 16 bytes if IPv6)
- SBV82 - Port number
  - Key/Length (2 bytes)

    X'8202'
  - Port (2 bytes)
- SBV85 - Host name of the client (SBV85 is not always present)
  - Key/Length (2 bytes)

    X'85*hh*', where *hh* is 1 + DNS name length
  - Flags (1 byte)

    80 - On if DNS name is truncated to 128 bytes
  - DNS name (Maximum 128 bytes)
- SBV86 - Zone ID if specified for a client IPv6 address (SBV86 is not always present)
  - Key/Length (2 bytes)

    X'8611'
  - Flags (1 byte)

    80 - On if zone is truncated
  - Zone ID (16 bytes)

Session initiation management (LOGAPPL, QINIT, FIRSTONLY, and DEFONLY)

The LOGAPPL, QINIT, FIRSTONLY, and DEFONLY options can be coded on DEFAULTAPPL, PRTDEFAULTAPPL, LINEMODEAPPL, LUMAP-DEFAPPL, or PRTMAP-DEFAPPL. For the remainder of this topic, DEFAULTAPPL represents all the default application statements.

- LOGAPPL, QINIT

  The LOGAPPL or QINIT functions keep the Telnet LU active if a Request Session fails due to the host VTAM application not being active. In addition, VTAM remembers the attempted Request Session and will initiate a session request to the Telnet LU on behalf of the application when the application becomes active. When the Request Session fails, Telnet sends the client a solicitor screen or USSMSG07

screen. The user then has the option of logging on to a different host VTAM application (if DEFONLY is not coded). When this different session is started, VTAM drops the queued Request Session for the original session.

What happens at session logoff depends on whether or not LUSESSIONPEND and FIRSTONLY are coded and whether LOGAPPL or QINIT is coded. If LUSESSIONPEND is coded, the connection remains; otherwise, it is dropped. If FIRSTONLY is coded, Telnet sends a USSMSG10 screen or solicitor screen to the client. If FIRSTONLY is not coded, Telnet will initiate another session with the default application defined by the DEFAULTAPPL statement. LOGAPPL and QINIT have different results when logging off the original application. When LOGAPPL is specified, a USSMSG10 or solicitor screen is sent to the client. When QINIT is specified, Telnet requests a session with the default application.

- FIRSTONLY, DEFONLY

  Sometimes a default application is used for the initial connection, but after LOGOFF a USSMSG10 or solicitor screen is more appropriate than redriving the default. In this case, code the FIRSTONLY parameter. This indicates the default should be used on the first session only. After a session has been established, any subsequent lookups ignore the default and send the USSMSG10 screen or solicitor screen.

  If MSG07, LUMAP-DEFAPPL, or PRTMAP-DEFAPPL is coded and the default application is inactive, an error screen will be sent to the client. The DEFONLY parameter will block a user-entered application choice if it is different than the default. This parameter prevents application choice while giving the user error information.

The following table summarizes several possible session initiation failure scenarios.

- ReqSess OK - A Request Session to the default application succeeded or a Request Session to a second application from the USSMSG07 screen succeeded.
- ReqSess Fail - A Request Session to the default application failed.
- 2nd Appl Fail - A Request Session to a second application from the USSMSG07 screen failed.

Scenario / Mapping Statement	ReqSess OK	ReqSess Fail		2nd Appl Fail	
		MSG07	No MSG07	MSG07	No MSG07
DEFAULTAPPL name	1	2	5	2	N/A
DEFAULTAPPL name LOGAPPL DEFAULTAPPL name QINIT	1	3	3	4	4

Figure 90. Session initiation failure scenarios

FIRSTONLY is not a consideration because the first session has not been established.

1. In session.
2. Send USSMSG07 or solicitor screen to client. Close the ACB.
3. Send USSMSG07 or solicitor screen to client. Keep ACB open, queue original session request in VTAM.
4. Send USSMSG07 or solicitor screen to client. Keep ACB open, keep the original queued session request in VTAM.
5. Drop connection.

The following table summarizes several possible session ending scenarios. The session is ending due to normal LOGOFF or session breakage (possibly caused by loss of the application).

- Original Application - User is in session with the original default application.
- CLSDST from Original Appl - User is in session with a second (or later) application after issuing a CLSDST-PASS from the original application.

Scenario / Mapping Statement	Original Application		2nd Appl or CLSDST from Orig	
	Logoff	Break	Logoff	Break
DEFAULTAPPL name DEFAULTAPPL name QINIT	1	1	1	1
DEFAULTAPPL name LOGAPPL	2	1	1	1
DEFAULTAPPL name FIRSTONLY LOGAPPL DEFAULTAPPL name FIRSTONLY QINIT	2	1	2	1

Figure 91. Session ending scenarios

In all cases, if LUSESSIONPEND is not coded the connection is dropped.

1. Request a session with the default application.
2. Send USSMSG10 or solicitor screen to client. Close the ACB.

Check client connection and connection/session takeover

A Telnet client can detect the loss of connectivity to Telnet without Telnet being aware of the loss. This condition is typically caused by a temporary problem in the network, such as a router experiencing a failure. In many cases, the network can correct itself and client retransmission can recover the data flow. However, the user often terminates the Telnet TCP connection at the emulator without waiting for the network to recover. In other cases, such as when a single router supports the client, the Telnet TCP connection is eventually terminated by the client TCP/IP stack after a certain number of retransmissions fail. When network connectivity is restored, the user initiates a new connection to Telnet. Even though the user is reconnected, in many cases the original SNA session cannot be associated with the new TCP connection because the session continues to be associated with the original TCP connection.

Assume that the host application is TSO and the user is in session with TSO user ID USER1. The route is lost and the user disconnects. The user then establishes a new Telnet connection without specifying an LU name. Telnet assigns a different LU to represent the client. When the logon to user ID USER1 is attempted, TSO fails the logon attempt because USER1 is still in session with the original Telnet LU. If the user does specify the same LU name when reconnecting, the connection is rejected sooner. Telnet fails the request during Early Lookup, indicating that the LU is already in use. The problem with both situations is that Telnet assumes that the original connection is still active and that the SNA session is still associated with that original connection.

The user can not terminate the original TCP connection and SNA session. The user has two choices:

- Wait for an inactivity timer in Telnet to clean up the original SNA session and close the original TCP connection
- Wait for an inactivity timer in TSO to initiate session termination

In the second case, either Telnet is configured to close the connection at session termination (NOLUSSIONPEND), or when Telnet attempts to refresh the TCP connection, Telnet detects that the TCP connection is gone and closes its representation of the connection. The end result in all cases is that the original SNA session and the original TCP connection are cleaned up after the specified inactive period of time has elapsed.

The user would like to quickly close the original connection to free the LU and the SNA session. If you specify the CheckClientConn parameter, Telnet checks the connectivity of all pre-existing connections associated with the client identifier of the new connection being established. The check is performed by sending a Timemark to each existing connection. The new connection is delayed early in the connection negotiation until all existing connections have responded or the specified wait time has elapsed. When all connections have responded, setup of the new connection continues. If some connections do not respond, then after the specified time elapses, Telnet closes any connections that did not receive a response. As soon as all of the unresponsive connections are closed, setup of the new connection

continues. The new connection is held until cleanup is performed to ensure that an LU is available and that the previous SNA session is cleaned up before the new connection continues negotiation. The CheckClientConn parameter is useful when multiple emulators exist on a single client and the user can tolerate the session being disconnected.

Rule: If you are using sysplex distributor to distribute connections across Telnet servers with CheckClientConn, you must use timed client affinity to ensure that the clients reconnect to the same Telnet server.

Tip: Be careful using the CheckClientConn parameter where proxy servers are being used and the Client Identifier is IP address rather than Host name or User ID. Telnet perceives all connections coming from the same IP address as coming from the same client. Depending on the number of connections, Telnet could send a large number of Timemarks every time a new emulator connects. A second parameter is available to limit the number of connections checked for a single client identifier.

Some users require that the same LU be assigned when the new connection takes over the old connection. The takeover function fixes this problem. When takeover is active for a connection, Telnet LU lookup attempts LU takeover upon entry to the lookup function, whenever a single LU name can be associated with a connection. Any number of existing Specific requests from one client identifier can be associated with new Specific requests. A single existing Generic request from one client identifier can be associated with a new Generic request. If there are multiple existing Generic requests, only the first is taken over and the remaining connections continue to hang.

The TKOSPECLU and TKOSPECLURECON statements activate takeover for the connection and require that the user specify the LU name. The statement name is derived from the function of connection takeover (TKO) for a Specific LU (SPECLU) connection request. Because the LU name is specified, LU lookup attempts to take over the original connection to which the LU was assigned. The Specific LU request allows a user to move to any client to take over a connection that is lost and provides some level of security because it requires the user to know the LU name.

The TKOGENLU and TKOGENLURECON statements activate takeover for a connection without the user specifying an LU name. The statement name is derived from the function of connection takeover (TKO) for a Generic LU (GENLU) connection request. During the original connection LU lookup, Telnet saves the LU name by Client Identifier. If another connection request is received from the same Client Identifier, Telnet assumes takeover should be attempted using the original LU name. If multiple connections are made from a single client, all additional connection setups will be delayed by the time it takes to attempt takeover of the first connection. When the takeover attempt fails, Telnet resumes normal Generic LU lookup for that connection. For Generic LU takeover to work, the Client Identifier must remain the same. The Client Identifier can be the client user ID derived from the client security certificate. If no user ID is available, the client hostname is used. If no hostname is available, the client IP address is used.

Rule: If you are using sysplex distributor to distribute connections across Telnet servers with TKOGENLU or TKOGENLURECON, you must use timed client affinity to ensure that the clients reconnect to the same Telnet server.

The TKOSPECLU and TKOGENLU statements cause the following events to occur. When the takeover connection request arrives, Telnet LU lookup discovers the LU name is in use and suspends the new request. Telnet sends a TIMEMARK request to the original client, which acts as an "are you there" message. The client is required to respond to the TIMEMARK. If no response is received by Telnet within the time specified on the takeover statement, Telnet drops the original session and connection. The original LU is reserved during the drop process. After the original session and connection are dropped, Telnet resumes processing the new request. This time the LU is not in use, only reserved for takeover purposes, and is assigned to the new takeover session. The user is essentially starting over. The original session has been dropped, allowing the user to immediately log on to the same TSO user ID again.

When sysplex distributor is used to distribute connections across Telnet servers with the TKOSPECLU or TKOSPECLURECON statements, the series of events can be extended to include the LUNS and possibly the previous owning LUNR. When the takeover connection request arrives at a LUNR and Telnet LU lookup discovers that the LU name is shared and is not already allocated to itself, the new request is suspended, and verification and allocation is requested from the LUNS. If the LUNS determines that the LU name has been allocated to another LUNR, the LUNS sends a verification request to the other LUNR. That LUNR sends a TIMEMARK to its original client. If the client responds, the owning LUNR informs the LUNS that

the LU name is still in use. If the client does not respond, the owning LUNR drops the original connection and tells the LUNS to deallocate the LU name. If the LUNS determines that the LU name is not in use, the name is allocated to the requesting LUNR. Depending on the response from the LUNS, the new LUNR then either accepts or rejects the takeover request.

The TKOSPECLURECON or TKOGENLURECON statements can be used to accomplish the same connection drop but avoid the session drop. When the original connection is dropped, the Telnet LU stays in session with the host application. The new connection is established and Telnet sends an LUSTAT to the host application indicating that Presentation Space Integrity was lost (X'082B'). Depending on the application, it will either end the session or resend the previous screen. By resending the previous screen, the user is able save the original session and avoid the SNA session tear-down and restart process. At worst, if the application drops the session upon receipt of the LUSTAT, the user is able to immediately log on again as if TKOSPECLU or TKOGENLU were coded.

In some cases, the original client TCP/IP stack may respond to the Timemark with a RESET. Telnet interprets this RESET as a client disconnect and assumes the user disconnected the session. Telnet then drops the session. To keep the session in this case, add the KeepOnTmReset option to the TKOSPECLURECON or TKOGENLURECON statement. A security risk exists when using this parameter. A user may actually disconnect just before the Timemark arrives due to an unauthorized takeover attempt. Telnet will interpret the disconnect as a response to the Timemark and allow the takeover without loss of the VTAM session.

Part of session initiation includes Telnet issuing a SETLOGON VTAM macro that contains control vectors, including control vector 64 described in “Connection information passed on the CINIT control vector 64” on page 634.

The control vectors include client information, such as IP address and host name if available. If the connection is capable of being taken over, a flag is set in the control vector. This information is passed to the application's logon exit for use by the application. If the connection is taken over, a second SETLOGON macro is issued. However, the control vector information does not go beyond the VTAM that processed the SETLOGON. The application's logon exit does not run again and is not aware of connection changes.

Some applications require the IP address to remain constant. Because the application cannot be notified of changes, Telnet has the ability to allow takeover by only clients from the same IP address. Specify SAMEIPADDR on the TKOGENLURECON or TKOSPECLURECON statements to ensure takeover is done by a client from the same IP address.

TKOSPECLU, TKOSPECLURECON, TKOGENLU, and TKOGENLURECON can be coded at all three parameter block levels for different levels of granularity. Code NOTKO to turn off all takeover function at any of the three levels.

The new connection must have an equal or higher security level to take over the original connection. The order for connection security is:

1. Basic
2. Secure
3. Secure with ClientAuthType Required
4. Secure with ClientAuthType SAFCHECK

If the original connection used SSL with ClientAuthType SAFCHECK, either takeover method will verify that the new connection is using a client certificate that maps to the same user ID. Telnet verifies this by translating the new certificate to a user ID and comparing the new user ID to the user ID on the original connection.

Tip: If you are using TKOSPECLURECON or TKOGENLURECON, you can specify SAMECONNTYPE to force the same connection type between the taker and the target connections. CV64 information is not updated when a session takeover occurs, and if a secure connection takes over a basic connection, the application is not updated with the new connection level because no new CV64 is sent to the application. You can also consider using takeover without reconnect (TKOSPECLU or TKOGENLU).

Guideline: If you map multiple certificates to the same user ID, a client presenting any of those certificates will be able to take over the connection. If there is a chance the connection can be taken

over by an unauthorized user, TKOSPECLURECON and TKOGENLURECON should not be used. Neither statement requires the user to reverify user authenticity to the host application.

Tip: A time value of zero is permitted. In this case the server will always perform the takeover whether or not the original connection is still active. The zero value is intended for testing purposes rather than production use.

Sometimes a takeover attempt will not complete as expected. This might be due to one of the following factors:

- The profile of the original connection defines how the original connection can be taken over. Be sure that the original connection supports the wanted takeover method.
- The new connection is not of the same connection type as the original session, and SAMECONNTYPE was specified on the TKOSPECLURECON or TKOGENLURECON statement.
- The new connection request must specify the LU name of the connection being taken over if TKOSPECLU or TKOSPECLURECON is specified.
- TKOSPECLURECON and TKOGENLURECON do not preserve a session if the takeover is done from a different port. The ACB of the LU is associated with the original port and must be closed before it can be associated with the new port. The takeover will function as a TKOSPECLU and TKOGENLU takeover.
- TKOSPECLURECON and TKOGENLURECON do not preserve a session if the takeover is done for a shared Telnet LU name managed by a LUNS.
- TKOSPECLURECON and TKOGENLURECON might not preserve a session if the original client connection is ended before the TKOSPECLURECON or TKOGENLURECON timer expires. If the close reason is TIMEMARK or INACTIVE, the session will be preserved under the assumption that the inactivity is due to a lost connection. Any other close reason will cause the takeover to function as a TKOSPECLU or TKOGENLU takeover. This is done to protect users who disconnect their client as a means of logging off their session. These sessions will not be taken over. Instead, the user will have to issue a new logon.

Takeover is also affected by where the new client is and how the old client responds. There are several event scenarios and results will vary.

- Event 1

  New client connects from a different IP address or Port.

  Original client responds to TIMEMARK.

  Result - Takeover will not occur in this case because the original client is still responding. With a Specific LU request, the new client will receive an error indicating the LU is already in use. With a Generic LU request, Telnet assigns the next available LU to the connection.

- Event 2

  New client connects from a different IP address or Port.

  Original client does not respond to TIMEMARK.

  Result - Connection takeover will occur. If TKOSPECLURECON or TKOGENLURECON is mapped to the client, session takeover will occur. A likely scenario in this case is Telnet has lost connectivity to the old connection due to a failed router and the new connection is using a different route, the original machine lost power and has not reestablished connectivity, or the original machine lost power but reestablished connectivity with a different IP address.

- Event 3

  New client connects from a different IP address or Port.

  Original client stack responds with RESET.

  Result - Connection takeover will occur. However, even if TKOSPECLURECON or TKOGENLURECON is coded, session takeover will not occur because Telnet handles the RESET as a client disconnect. A likely scenario in this case is a PC lost power and then regained power. The takeover request is accepted from either a different PC or the same PC using a different port. After the new connection request is accepted, Telnet sends a TIMEMARK to the original client PC stack. The PC stack does not recognize the

IP-port and responds with a RESET. If KeepOnTmReset is specified and the RESET is received by Telnet after the Timemark has been sent, Telnet will keep the session.

- Event 4

  New client connects from the same IP address and Port.

  Telnet stack rejects the request.

  Original client times out and sends a RESET.

  Result - The original session and connection are dropped. Takeover does not occur. The new client is able to connect on retry because the original connection and session were cleaned up. A likely scenario in this case is a PC has lost power and then regained power. The same PC is used to attempt takeover. The client is assigned the same port as before the power loss and has the same IP address.

Queueing sessions

Logon manager applications are very popular. Typically they are set up as a default application which sends a selection screen to the user. After the user specifies the destination application choice, the logon manager typically issues a CLSDST macro with OPTCD=PASS to the destination application. A new session is started with the destination application. The logon manager session is closed with a special UNBIND sent to Telnet indicating that a new session BIND is forthcoming. Telnet receives that special UNBIND and then waits for the next BIND instead of cleaning up as it would when receiving a normal UNBIND. When the user logs off the destination application, Telnet either goes through the initial database lookup process again (which will result in a session with the logon manager) or drops the connection, depending on whether LUSESSIONPEND is coded. Logoff of the original application will cause Telnet to perform normal close function instead of leaving the LU ACB open.

Many logon managers were written to support real terminals, not Telnet, and issue a SIMLOGON OPTCD=Q immediately after issuing the CLSDST-PASS. When SIMLOGON Q is coded, the logon request is added to a VTAM queue. The first application queued is the first application off the VTAM queue. Immediately after user logoff from the destination application, VTAM (on behalf of the logon manager) will request a session with the terminal LU (or Telnet LU representing the client). This works very well for real terminals, but causes timing problems for Telnet when the logon manager is a default application. In this case, Telnet and VTAM both end up requesting a session.

The QSESSION option on ALLOWAPPL or RESTRICTAPPL can be used to correct this timing problem. When coded for the logon manager, Telnet will not do normal close processing when the UNBIND from the destination application arrives. Telnet will leave the LU ACB open and wait for the BIND from the logon manager that is generated because of the Queued SIMLOGON. When the BIND does arrive, Telnet will verify that the application name is the original logon manager and finish session setup.

If QSESSION is specified for an application that does not queue a SIMLOGON, or the SIMLOGON is removed from the queue because the original application was recycled, Telnet will be waiting for a BIND that is never coming. The connection will appear to be hung. To safeguard against this hang condition, a timer can be started when the destination application UNBIND is received. Telnet will wait for the specified period of time for a BIND from the QSESSION application. If the timer expires and there is no session, Telnet will clean up the connection as if the QSESSION parameter had not been specified.

As an example of the QSESSION parameter, assume that APPL1, APPL2, and APPL3 are each defined in VTAM. APPL1 will issue a SIMLOGON-Q after CLSDST-PASS. The following Telnet statements allow connections to access the applications and define which is a QSESSION application.

```
ALLOWAPPL APPL1 QSESSION,3
ALLOWAPPL APPL*
```

The client first logs on to APPL1. APPL1 issues a CLSDST-PASS to APPL2 and a SIMLOGON-Q. Finally, APPL2 issues a CLSDST-PASS to APPL3. When the APPL3 session is ended, VTAM sends an APPL1 BIND to Telnet. If the queued SIMLOGON had been removed, Telnet would have continued cleanup 3 seconds after receiving the UNBIND from APPL3. When the APPL1 session is ended, the ACB is closed and Telnet either goes through the initial database lookup again or closes the connection, depending on whether LUSESSIONPEND is coded.

As a second example, assume that APPL2 also issues a SIMLOGON-Q after it issues a CLSDST-PASS to APPL3. As in the previous example, when the APPL3 session is ended, VTAM sends an APPL1 BIND to Telnet. If the queued SIMLOGON had been removed, Telnet would have continued cleanup 3 seconds after receiving the UNBIND from APPL3. When the APPL1 session is ended, VTAM sends an APPL2 BIND to Telnet. The APPL2 SIMLOGON-Q was queued behind the APPL1 SIMLOGON-Q in VTAM.

Disconnect on session error

The DISCONNECTABLE option on either the ALLOWAPPL or RESTRICTAPPL statement determines what type of session termination to send to the host VTAM application when Telnet initiates session termination. If DISCONNECTABLE is coded, Telnet issues a TERMSESS UNBIND(0F). Otherwise, Telnet issues a TERMSESS UNCOND. For example, when DISCONNECTABLE is coded for the TSO application, an unexpected connection loss results in an UNBIND(0F) being sent to TSO putting it in a reconnectable state. The DISCONNECTABLE parameter has no effect on a session ended normally by the user logging off the session. The QSESSION parameter can be coded with DISCONNECTABLE on either statement.

Bypass RESTRICTAPPL with CERTAUTH

CERTAUTH is an option on RESTRICTAPPL used in conjunction with client authenticated secure connections or Express Logon. In both cases the client certificate is used to derive a user ID. If the user chooses an application that is a RESTRICTAPPL, the normal Telnet response is to request a valid user ID and password before allowing access to the application. However, if the user has been authenticated with a client certificate it may not be necessary to require a user ID and password. With the CERTAUTH option on RESTRICTAPPL Telnet will use the derived user ID. If the user ID is valid (listed on the RESTRICTAPPL statement), Telnet will bypass the user solicitation and immediately give access to the application. The derived user ID value depends on the type of connection. If Express Logon is being used, the user ID is derived from the latest Client Certificate/Applid combination received from the client. If Express Logon is not being used, the user ID is the Client Identifier user ID derived from the Client Certificate from the SSL handshake.

Allow printer sessions with RESTRICTAPPL

The ALLOWPRINTER option enables printers to establish a session with an application defined as a RESTRICTAPPL. Telnet verifies all session requests before finishing session setup, and if the application name is restricted by the RESTRICTAPPL statement, Telnet requires the user to enter a user ID and password before session initiation begins. The user of a printer emulator has no way to provide a user ID and password, and you might want to allow printer sessions without the user ID and password verification, mimicking the behavior of the ALLOWAPPL statement. The ALLOWPRINTER option on RESTRICTAPPL gives you the flexibility to keep the terminal LU sessions restricted while allowing all printer LU sessions access to the application, without user ID and password. In most cases, printer sessions originate from the application and not the user, keeping security control in the application. However, use this option with care if you have coded PRTDEFAULTAPPL or PRTMAP-DEFAPPL. In this case, it is possible for users to originate a printer session.

Keeping the ACB open

Some host VTAM applications are set up to inquire whether a secondary LU is active. If the LU is active, the application initiates a session. The LUMAP option, KEEPOPEN, causes the LU to be activated when the LU is assigned to the connection, and to remain active for as long as the LU is assigned to the client by this mapping statement.

LU assignment is different for TN3270E and TN3270 connections. TN3270E connections have an LU assigned during connection negotiation. TN3270 connections do not have an LU assigned until the VTAM application name is known.

When the host VTAM application initiates a session with the secondary LU, the host must issue an INQUIRE to see if the LU is active with an OPEN ACB. This INQUIRE will fail for Telnet LUs because Telnet does not open the ACB until a session request is sent from Telnet to VTAM. When the user gets the solicitor (or USSMSG10) screen, Telnet has not opened the ACB of the LU assigned to a TN3270E connection. TN3270 and LineMode connections do not have LUs assigned yet. If the KEEPOPEN

parameter is coded on the LUMAP statement used by Telnet to assign an LU during Early Lookup for a terminal TN3270E connection, Telnet opens the ACB before sending the solicitor (or USSMSG10) screen. At that time a user can either log on to an application as usual or wait for a host application to INQUIRE about the LU and initiate a session. TN3270 connections will not have an LU assigned until an application name is known. When the name is known, an LU is assigned to the connection, the ACB of the LU is opened, and a session request is issued. If the KEEPOPEN parameter is coded on the LUMAP statement used by Telnet to assign the LU, the LU stays assigned to the connection with the ACB open until the connection is dropped. If profile statements define LUs uniquely to different applications, a second logon to a different application might fail. When KEEPOPEN is mapped to a connection, the MSG07 and LUSESSIONPEND functions are in effect whether or not they were explicitly coded. When a session is ended, the connection remains and the ACB remains open. Only a client disconnect, a Telnet error, or the KEEPINACTIVE/INACTIVE timers will cause a KEEPOPEN connection to be dropped. The KEEPINACTIVE timer is used whenever the Telnet LU is not in session with a VTAM application. Otherwise, the INACTIVE timer is used. See "Connection persistence" on page 604 for information about ending idle KEEPOPEN connections.

Express Logon Feature

The Express Logon Feature (ELF) allows a user to connect to an MVS host VTAM application without explicitly entering a user ID or password. Telnet uses the client certificate to resolve the user ID and RACF generates a Multi-Factor Authentication (MFA) token or a temporary password called a PassTicket. ELF requires a secure connection with level 2 client authentication, a client that supports ELF, and RACF MFA or PassTicket setup.

The ELF function is activated by specifying the EXPRESSLOGONMFA parameter (to use MFA) or the EXPRESSLOGON parameter (to use PassTickets). The function can be deactivated by specifying NOEXPRESSLOGONMFA or NOEXPRESSLOGON. Any of these parameters can be coded in TELNETGLOBALS, TELNETPARMS, or PARMSGROUP. For a detailed discussion of ELF, see Appendix C, "Express Logon Feature," on page 1335.

Device types and logmode considerations

The VTAM logmode defines many characteristics of the session established between the Telnet LU representing the client and the host VTAM application. For example, the logmode defines response types, presentation style, and the type of LU Telnet is emulating. LU0 (non-SNA) and LU2 (SNA) represent terminal LU types. LU1 (SCS) and LU3 (3270 Data) represent printer LU types.

Telnet matches a VTAM logmode to each client as it connects based on the client device type, unless the user specifies a logmode on the USSMSG10 screen or logmode is configured in the USS table mapped to the connection. See z/OS Communications Server: IP Configuration Reference for default device type and logmode table information. The default terminal logmodes are non-SNA for TN3270 connections and SNA for TN3270E connections. At session request time, Telnet indicates to VTAM the appropriate logmode based on device type. The host VTAM application usually honors the request and binds the session using the requested logmode. However, depending on VTAM statements, the application can override the requested logmode and bind the session using different characteristics than Telnet requested. For this reason, some screen sizes might not work correctly even though the logmode defined in Telnet is correct. If the KEEPOPEN function is used to allow session initiation by the host application, the wanted logmode must be coded on the DLOGMOD parameter as part of the VTAM application definition statement that defines the Telnet LU. Otherwise, the application will choose its own logmode.

Telnet processes the ATTN KEY request differently for non-SNA and SNA sessions. For non-SNA sessions (BIND FM value 02), Telnet converts the ATTN KEY request to a '6C'x data byte and sends it to the application. For SNA sessions (BIND FM value 03), Telnet converts the ATTN KEY request into a SNA signal and sends it to the application as expedited data. Some clients send both an ATTN KEY function code and a '6C'x data byte to ensure the ATTN is seen by the application. Telnet converts the ATTN KEY function into either a '6C'x data byte or a SNA signal and also forwards the '6C'x data. Some applications give unexpected results or Telnet might appear to not support ATTN when two ATTNs are received. The SINGLEATTN parameter causes Telnet to drop the second ATTN if it immediately follows an ATTN. The SINGLEATTN and NOSINGLEATTN parameters can be coded at all three parameter block levels for different levels of granularity.

To change either the TN3270 or the TN3270E logmode for a device type, use the TELNETDEVICE parameter. Whenever Telnet initiates the session request, Telnet will request that the logmode specified on the TELNETDEVICE statement be used for the session. The application (the primary LU) does have the ability to override the requested logmode and use a completely different logmode. The TELNETDEVICE parameter can be coded in all three parameter block levels for different levels of granularity. Coding TELNETDEVICE in a PARMSGROUP that is mapped on the LUMAP-DEFAPPL-PMAP statement enables the logmode to be LU and application specific.

If the application initiates the session, the TELNETDEVICE logmode has no affect on the session. For example, printer sessions are initiated by the application unless a printer default application is specified. The LUMAP-KEEPOPEN parameter can be used to open a terminal ACB and wait for the primary application to initiate the session.

Transform Linemode connections can have a unique logmode by coding TELNETDEVICE with a device type of TRANSFORM. Any logmode used must not support extended graphics.

The special case logmode NONE can be specified indicating that Telnet should not send any logmode request when initiating the session.

In the examples that follow, the first line causes only the TN3270 logmode to change from the default to SNX32705. The second line causes both the TN3270 and TN3270E logmodes to change from their defaults to SNX32705 and SNX32702. The third line causes only the TN3270E logmode to change from the default to SNX32702.

```
TELNETDEVICE 3278-5-E SNX32705
TELNETDEVICE 3278-5-E SNX32705,SNX32702
TELNETDEVICE 3278-5-E ,SNX32702
```

Using the Telnet solicitor or USS logon screen

This topic describes the Telnet solicitor screen and Telnet Unformatted System Services (USS) support. You can use the Telnet solicitor screen to obtain all information that is needed to establish a session. However, Telnet is often the primary method of connecting to the SNA mainframe environment; SNA users are accustomed to entering abbreviated logon commands, specifying their own logmode, and entering user data from SNA terminals. The USS table itself can be used to customize information such as logmode before it is sent to VTAM. The logmode specified by the user or the USS table overrides the logmode specified by the TELNETDEVICE statement. For ease of migration, Telnet simulates SNA USS processing very closely. This simulation extends to being able to use the same assembled USS tables that are used by VTAM. VTAM-only character substitutions are ignored by Telnet and Telnet-only character substitutions are ignored by VTAM. Blanks are used in their place. To further extend the simulation of SNA terminals, Telnet also supports all of the INTERPRET table function.

Using the Telnet solicitor logon screen

Telnet sends a solicitor screen to the user if one of the following conditions is true:

- No DEFAULTAPPL, LINEMODEAPPL, USSTCP, or LUMAP-DEFAPPL mappings match the client's Client Identifier.
- The requested application is a RESTRICTAPPL.

The following example shows the Telnet solicitor screen:

```
Enter Your Userid:
Password:                          New Password:
Application:
```

Initial cursor placement can be specified. Where initial placement should be depends on client macros used and user preferences. The OLDSOLICITOR parameter is used to implement this choice. The default cursor position is on the 'Application:' field. If OLDSOLICITOR is coded, the cursor is positioned on the 'Enter Your Userid:' field. The OLDSOLICITOR and NOOLDSOLICITOR parameters can be coded at all three parameter block levels for different levels of granularity.

Use the PASSWORDPHRASE parameter to configure Telnet so that the solicitor screen can accept either a password or password phrase. The screen layout changes to accommodate the potentially larger password phrase length. The following example shows the Telnet solicitor screen when the PASSWORDPHRASE parameter is specified:

```
Enter Your Userid:
Password:

New Password:

Application:
```

In addition to satisfying RESTRICTAPPL, there are other times when a user might want to use the user ID and password fields. For example, the solicitor screen might also be used to change a password by entering the user ID, old password, and new password. The application field does not need to be filled in. If insufficient information was provided by the client (for example, a user ID but no password), the Telnet solicitor screen is returned with a message prompting for the required field. A message is also returned if the security program encounters an error when attempting to change the password. Example messages include:

- Password required
- Password is not authorized

Using the Telnet USS and INTERPRET support

The Telnet USS function provides the user with a USSMSG10 logon screen similar to the logon screen used by native SNA terminals. The Telnet USS function supports sending USSMSGs to the client, receiving and parsing USSCMDs from the client, and using a translation table defined in the USS table.

Telnet supports both 3270-format and SNA character stream (SCS)-format USS tables. The SCS USS table name is optional on the USSTCP mapping statement. If an SCS name is provided, the SCS USS table is used for all TN3270E connections.

The CLEAR key is handled differently by the client depending on whether a 3270-format or SCS-format USS table is used. The CLEAR key has a special function for 3270-format USSMSGs. Pressing the CLEAR key sends the CLEAR key data stream to Telnet. If REFRESHMSG10 is specified or used by default, pressing the CLEAR key refreshes the screen with USSMSG10 for any USSMSG other than USSMSG10. If you press the CLEAR key while at USSMSG10, the screen is cleared and the cursor is placed near the upper left corner (row 1, column 2). If you press the CLEAR key a second time, the screen refreshes with USSMSG10. If NOREFRESHMSG10 is specified, pressing the CLEAR key always clears the screen of any USSMSG and places the cursor in the upper left corner (row 1, column 1). For SCS-format USSMSGs, the CLEAR key does not provide the same special function. For SCS, the client processes the CLEAR key by clearing the screen and placing the cursor at the upper left corner (row 1, column 1) with no data sent to Telnet.

USS data traffic is also handled differently depending on whether a 3270-format or SCS-format USS table is used. If a 3270-format USS table is used with a TN3270E connection that has negotiated BIND-IMAGE, a Telnet-generated Bind/Unbind encapsulates the USS traffic. If the connection is TN3270, BIND-IMAGE is not negotiated, or the SCS-format USS table is used, Telnet does not encapsulate the USS traffic in a Bind/Unbind. If a 3270-format USS table is used, the Telnet data type for all traffic is 3270-DATA. If an SCS-format USS table is used, the Telnet data type for all traffic is SSCP-LU-DATA. Regardless of USS table format, if the SYSREQ function is supported on a TN3270E connection and SYSREQ is received by Telnet, Telnet will accept a LOGOFF command in SSCP-LU-DATA format. If the client sends any command other than LOGOFF, `COMMAND UNRECOGNIZED` is returned in SSCP-LU-DATA format. If the client sends a second SYSREQ, Telnet reverts back to whatever the session state was prior to receiving the first SYSREQ.

USSCMD parsing also includes checking for INTERPRET table entries that might provide more function than USS tables alone can provide. Sample USS tables are in TCP/IP data sets SEZAINST(EZBTPUST) and SEZAINST(EZBTPSCS). A sample INTERPRET table is in TCP/IP data set SEZAINST(EZBTPINT). The 3270 format USS sample has been assembled, linked, and loaded into the product data set. The tables can be used by coding the USSTCP and INTERPTCP mapping statements in BEGINVTAM. For example, the

statements below will map the sample tables to the client at IP address 1.1.1.1. See “Mapping Objects to Client Identifiers” on page 607 for mapping details.

```
USSTCP      EZBTPUST,EZBTPSCS  1.1.1.1
INTERPTCP   EZBTPINT  1.1.1.1
```

A new table can be created at any time and link-edited. Customized USS and INTERPRET tables can be created to change messages, commands, and translation tables. For example, messages can be changed to have non-English text or to have different syntax. Commands can be changed to accept different syntax or to have different default values. A VARY TCPIP,*tnproc*,OBEYFILE command will cause Telnet to load the new table with the new profile being processed. Any new connection using the new profile will be assigned the new table. Telnet also supports dynamic updating of same-name USS or INTERPRET tables. The VARY TCPIP,*tnproc*,OBEYFILE command adds the new version of the table to the new profile. New connections use the new copy associated with the new profile while old connections continue to use the old copy associated with the old profile.

USS table customization

Customized USS tables are used by both VTAM and Telnet, with any product-specific character substitutions converted to blanks. For example, @@SSCPNM is blank for Telnet and @@PRT is blank for VTAM. Telnet USS processing also supports system symbolic substitution. VTAM does no substitution for system symbolics. The tables must be in a data set that is in the system's linklist or is in the STEPLIB statement of the TCP/IP startup procedure. Any changes to a Telnet USS table should be made with supplementary user-defined USS tables. The IBM-supplied USS table should not be changed as it provides a good example of coding most commands and messages. Telnet loads the first table that is found with the name EZBTPUST and defines it as the default USS table. If this table is not found, there is no default USS table. Whether to include a default USS table depends on the wanted message output. When writing a USS Message, Telnet searches the USS table that is mapped to the client first. If the message does not exist in the mapped table, Telnet searches the default table. If the message does not exist in the default table, Telnet writes USSMSG14. If no default table exists, Telnet generates a USSMSG14. For 3270 format USSMSGs, the user can get back to the USSMSG10 from any message by pressing the CLEAR key. The default table does not affect the USS commands. The command entered must be in the mapped table or it is not recognized. The sample SCS USS table found in SEZAINST(EZBTPSCS) is not assembled and linked, and it is not loaded into Telnet as a default SCS USS table. The table must be assembled, linked, loaded, and mapped in the Telnet profile to be used.

Creating a USS table

The following macro instructions are used to create the USS table. Telnet USS function supports almost all VTAM session-level USS message and command definitions. See z/OS Communications Server: IP Configuration Reference for macro details.

- USSTAB indicates the beginning of the USS table.
- USSCMD defines commands accepted by Telnet.
- USSPARM defines each operand or positional parameter that can be specified on the USSCMD macro instruction. It also defines default values for the operand or positional parameter. Multiple USSPARM macro instructions can be associated with a USSCMD macro instruction. For each operand or positional parameter code a USSPARM macro instruction.
- USSMSG defines messages sent from Telnet.
- USSEND indicates the end of the USS table.

The following rules are some of the more common rules to consider when you are coding a new USS table. Also, see the samples in SEZAINST(EZBTPUST) and SEZAINST(EZBTPSCS) as a guide. “Considerations when using mixed-case passwords” on page 647 describes general table rules.

- If a DEFAULTAPPL application is mapped at the same Client Identifier level as a USS table, or an LUMAP-DEFAPPL application is mapped, the USS table is used only to return error messages and optionally after the first session logoff. FIRSTONLY or LOGAPPL options on DEFAULTAPPL will cause Telnet to send a USSMSG10 after the first session logoff. DEFAULTAPPL without the FIRSTONLY or LOGAPPL options will cause Telnet to request a session with the default application after every session logoff.

- Both the 3270 data stream and the SNA character stream (SCS) formats are supported. For more information, see *3270 Data Stream Programmer's Reference* and the table samples.
- If a user-defined table is coded as part of another module, code an assembler EXTRN definition statement for the table name in that module so the table will be known externally and can be accessed by other modules.

Below are message related rules.

- 3270 format USSMSGs must contain the 3270 data stream write control characters (WCCs).
- All character substitutions (@@'s) substitute the same number of characters. Any character substitution that is VTAM-specific will be translated to blanks. If the substituted value is smaller, the field is padded to the right with blanks. The parameter LUNAME or SCAN must be coded on the USSMSG macro instruction for Telnet to perform character substitutions. For a complete list of character substitutions, see z/OS Communications Server: IP Configuration Reference for Telnet and z/OS Communications Server: SNA Resource Definition Reference for VTAM. Telnet supports multiple USSPARMs with the DATA keyword. This method can be used to pass multiple data parameters to the host application. For example, two DATA USSPARMs allow the user to type 'TSO USER1 PROC001' and have both the user ID and the Procname passed to TSO as data. Telnet also supports the system symbolics substitution, padding to the right when the substituted value is smaller than the symbolic. VTAM USS processing does not support system symbolic substitution.

Below are command related rules.

- LOGON command format

  PL1 - logon applid(tso) logmode(snx32702) data(user1)

  BAL - logon applid=tso,logmode=snx32702,data=user1
- Any application defined in a USSCMD macro instruction must also be specified on either an ALLOWAPPL or a RESTRICTAPPL statement in the Telnet profile.
- If the USS Command rules in z/OS Communications Server: IP Configuration Reference cannot be followed, use an interpret table to convert the character-coded command into a formatted SNA request.

Considerations when using mixed-case passwords

Mixed-case passwords can be used by applications if an SAF-compliant security product (such as RACF) has enabled this support. In some cases, the USS LOGON DATA parameter is used to send the password to the application. If a terminal user enters a mixed-case password on the USS LOGON command and it is translated to uppercase by the translation table, the logon will fail if the target application expects the password in mixed case.

The USS LOGON command is displayed on the terminal as it is typed, so the password is displayed unless the 3270 format is used and the password is entered into a field with a non-display attribute. For additional security, inform the terminal user to stop entering the password as part of the USS LOGON. Instead, the application should prompt the terminal user for the password in a non-displayed field. If mixed-case passwords are used and the terminal user continues to enter the password as part of the USS LOGON command, the logon will fail when using TRANSLATE=YES (the default) on the USSPARM, because the password has been translated to uppercase.

If you want to continue allowing the terminal user to enter the password on the LOGON command, use one of the following methods to support mixed-case passwords:

- If the current USS translate table is used to set all characters to uppercase, the TRANSLATE=NO operand can be added to the USSPARM macros for the corresponding USSCMD, to prevent the specified DATA USSPARM containing the password from being translated to uppercase. For details, see z/OS Communications Server: IP Configuration Reference.
- Do not change the USS command and inform the terminal user to enter the DATA portion of the USS LOGON in single quotes. USS will not translate data within single quotes, and the quotes are removed before the data is passed to the application.
- If the terminal user is specifying only APPLID and DATA on the USS LOGON command, you can use an interpret table. The Telnet Interpret function passes all entered data without translation to the

application. Specify REMOVE=Y on the LOGCHAR macro to remove the first non blank string from the entered data. Because the terminal user could enter the APPLID in lowercase, you should set up an interpret table that searches for both an uppercase and a lowercase APPLID.

With each of these methods, if the user ID is entered with the password, you must first verify whether the application supports translating the user ID to uppercase. A simple test is to enter the DATA portion of the USS LOGON in single quotes with the user ID specified in lowercase. USS will not translate data within single quotes and the quotes are removed before the data is passed to the application. If the logon fails, the application does not support translating the user ID to uppercase and the terminal user must enter the user ID in uppercase and the password in mixed case for the methods suggested.

INTERPRET table customization

The standard Telnet USS logon support should meet the needs of most installations. However, Telnet does support INTERPRET table function if special circumstances require accepting a sequence of characters outside the normal USS command format. For example, the user might want to enter logon data that includes blanks. The INTERPRET table defines all entered data, including blanks, as a USSPARM DATA entry. The PL1 USSCMD format treats each blank as a parameter delimiter and cannot properly process a variable number of parameters. The INTERPRET table character sequences are scanned whenever the client is mapped to both a USS table and an INTERPRET table. Both must be mapped because the INTERPRET function is a subset of the USS function. INTERPRET is not a stand-alone function. The sample INTERPRET table found in SEZAINST(EZBTPINT) is not assembled and linked, and it is not loaded into Telnet as a default INTERPRET table. The table must be assembled, linked, loaded, and mapped in the Telnet profile to be used.

Creating an INTERPRET table

Telnet INTERPRET function supports all functions provided by the VTAM INTERPRET definitions. See z/OS Communications Server: IP Configuration Reference for macro details. The following macro instructions are used to create an INTERPRET table:

- INTAB indicates the beginning of the INTERPRET table.
- LOGCHAR defines a single logon message and name of an application program.
- ENDINTAB indicates the end of the INTERPRET table.

Below are some of the more common rules to consider when coding a new INTERPRET table. Also, see the sample found in SEZAINST(EZBTPINT) as a guide.

- The LOGCHAR APPLID= supports APPLICID, ROUTINE and USERVAR.
- Code the most restrictive, or longest, LOGCHAR SEQNCE values first. Otherwise, unexpected matches can occur. The table is scanned from top to bottom until a match is found whether or not it is the most exact match. For example, assume sequence 'LOGA' is assigned APPL1 and any other 'LOG' sequence is assigned APPL2. If sequence 'LOG' is before 'LOGA', entry 'LOGA' will never be found even when the user enters 'LOGA' because entry 'LOG' will be the first match. All sessions will go to APPL2. The problem is corrected by putting 'LOGA' before 'LOG' in the table.

Assemble, link, and load a table

Use the sample JCL in SEZAINST(EZBUSJCL). In the sample, the USS table is in USER1.TABLES(USSTEST). It must be assembled and link-edited into the system's linklist or into a library concatenated as a STEPLIB in the TCP/IP startup procedure. In the sample, the table is link-edited into USER1.LINKLIB(USSTEST). The same procedure can be used for the INTERPRET table. Change the name of the input file source and the link-edit target member. The VTAM USS and INTERPRET macros used for the assemble can be found in *hlq*.SISTMAC1.

SMF

The Telnet server writes the following SMF records:

- “Client activity records” on page 649
- “Profile configuration records” on page 649

Network management applications can send these records to a TCP/IP stack for NETMONITOR processing. For more information, see "Real-time SMF information service access control" on page 147.

Client activity records

SMF records are written when a user establishes a session (SMF LOGN or Telnet SNA Session Initiation record) and when the session is ended (SMF LOGF or Telnet SNA Session Termination record). Optional SMF recording is controlled by using the SMFINIT and SMFTERM statements.

Two different record formats are available: SMF type 118 and 119. The type 119 records were first introduced in z/OS V1R2 Communications Server, and are controlled by use of the TYPE119/NOTYPE119 operands on the SMFINIT and SMFTERM statements. The subtypes cannot be changed for type 119 records and are set to the STD values. The use of the STD operand or the specification of a nonstandard subtype number on the SMFINIT and SMFTERM parameters control the usage of the older type 118 record processing. Data that are recorded include the application name, Telnet LU name, client and host IP address and port, time of logon or logoff, and data count in and out. Combined with the SMF utility exit routine, SMF data can be used to track Telnet usage by a number of variables. If statements for both format types are coded, then both record types are written. Use that capability sparingly because of the additional processing cost that is involved in generating both types of records. For more information about the layouts for type 118 and type 119 SMF records, see z/OS Communications Server: IP Programmer's Guide and Reference.

Profile configuration records

You can control the writing of SMF type 119 Telnet profile configuration records by using the SMFCONFIG statement in the TELNETGLOBALS statement block. The data that is recorded in this SMF record includes the following data:

- Profile data sets used
- Options and values in the TELNETGLOBALS, TELNETPARMS, and BEGINVTAM blocks
- LU mapping statements
- Client identification statements

For more information about the type 119 Telnet profile configuration records, see z/OS Communications Server: IP Programmer's Guide and Reference.

Connection monitoring mapping statement

Telnet collects monitoring data for any client connection that is mapped to a MONITORGROUP Object. The collected data can be retrieved several ways:

- Display the data for a single connection using the D TCPIP,*tnproc*,TELNET,CONN,CONN=*connid* command.
- A network management agent can request the data from the Telnet SNMP subagent, after the subagent is started by specifying the TNSACONFIG statement. TCPIPJOBNAME must be specified for the subagent to connect with the agent. For information about the TNSACONFIG statement and the parameters needed to start the subagent, see z/OS Communications Server: IP Configuration Reference. For details about configuring the SNMP agent, see Chapter 23, "Simple Network Management Protocol," on page 1237.
- The Network Management Interface (NMI) Application Programming Interface (API), EZBNMIFR, can be used instead of SNMP to retrieve the data. In this case, the TNSACONFIG statement is not needed. For details about configuring EZBNMIFR and the parameters needed to issue the call, see z/OS Communications Server: IP Programmer's Guide and Reference.
- The Telnet SMF119 session termination record contains two optional sections that can be used to retrieve life-of-session performance data. Life-of-session data is a subset of life-of-connection data. Sliding-window data is not available because data is reported only at the end of the session and not at periodic intervals. One optional section contains response time statistics (excluding sliding-window) and a second optional section contains response time counts by time bucket. For details on the Telnet

server SNA session termination record, see z/OS Communications Server: IP Programmer's Guide and Reference.

Telnet is capable of collecting two types of response time data at the connection level.

- Response time statistics
  - Life-of-connection response time averages
  - Sliding-window response time averages
  - Sum of squares for variance and standard deviation calculations
- Response time counts by time bucket

The MONITORGROUP Object statement in BEGINVTAM is used to set the criteria for monitoring. Options allow the inclusion or exclusion of averages, counts by time bucket, and whether or not the IP network should be included in measurements. When a MONITORGROUP is mapped to Client Identifiers with the MONITORMAP mapping statement, the requested data will be collected for those connections.

Up to 255 MonitorGroups can be active at one time for all ports per Telnet instance. When 255 MonitorGroups are active, no additional groups can be created until some groups are no longer in use by Telnet. A MonitorGroup is considered no longer in use when the group is not defined in a current profile and no connections are using the non-current profile where the MonitorGroup was defined. After being notified that no entries are available, it is up to the operator to manage the removal of some entries by ending connections to free a profile. The operator does not get a notification when an entry becomes available.

Each MonitorGroup entry in the table is assigned an index number. The index number coincides with one of the 255 slots available in the MonitorGroup table. Each Telnet connection entry saves the index number of the MonitorGroup mapped to the connection. When the management application queries a connection for data, it can also query the MonitorGroup table to get the group name and configuration values based on the index found in the connection entry. With the data from the connection and the MonitorGroup criteria from the MonitorGroup table, a management application can generate several summary reports. If a management application is not being used to collect the connection monitoring data, the D TCPIP,*tnproc*,TELNET,CONN,CONN=*connid* command can be used to view all the collected data and MonitorGroup information for a single connection.

Collecting response time data

The typical Telnet data flow can be represented as shown in Figure 92 on page 650.

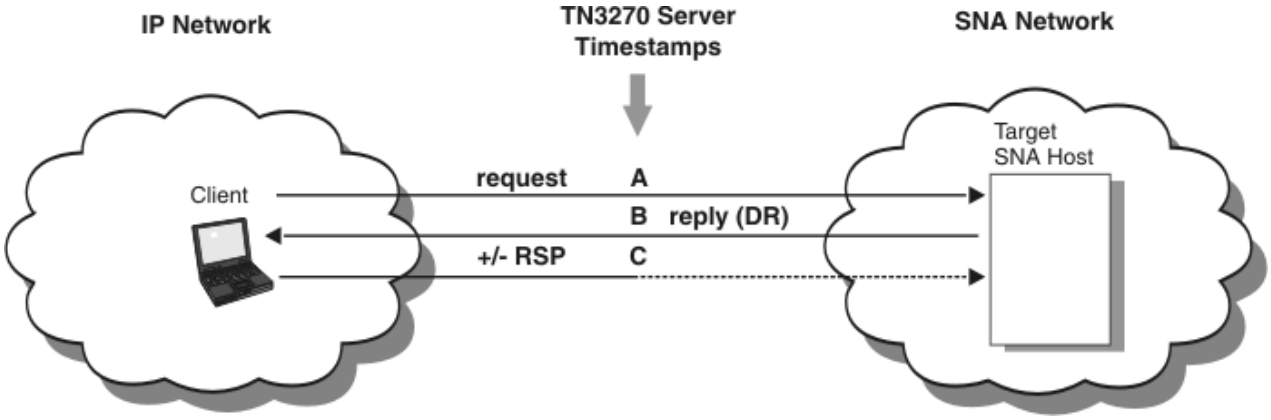


Figure 92. Typical Telnet data flow

Telnet saves timestamp A when it sends a client data request to VTAM, saves timestamp B when the target SNA host application returns data, and saves timestamp C when the client responds to the definite response request that flowed with the reply.

There are many clients, mostly those not supporting TN3270E negotiation, that do not support the Definite Response (DR) function. In this case, Telnet approximates IP response time by appending a TIMEMARK request immediately after the data. A Telnet TIMEMARK acts as a synchronization mark or as

an "are you there" function for almost all clients. Almost all clients respond to the TIMEMARK request. Timestamps B and C are set based on when the TIMEMARK request is sent and a TIMEMARK response is received. In rare cases, a client does not respond to a TIMEMARK request. If Telnet does not receive a response to the TIMEMARK, a flag in the connection data indicates that IP transit time measurements were not attempted.

Some installations might not want to incur the additional network traffic of DRs or TIMEMARKs to measure IP transit time, and are interested only in the SNA application side transit time. In this case, turn off IP response measurements by specifying NOINCLUDEIP within MonitorGroup. Total response times will be the SNA response time only.

Specify INCLUDEIP within the MonitorGroup to include IP response measurements when monitoring a connection. It does not matter whether the SNA host application requested a definite response on its reply. Many applications save processing time and IP transactions by not requesting Definite Response. If IP response time monitoring is wanted and the client supports DR, specify DYNAMICDR within MonitorGroup to add the definite response request to ensure that a response is received from the client. In this case, Telnet does not forward the response to the target SNA host application, as indicated by the dotted line in Figure 92 on page 650. Specify NODYNAMICDR to turn off dynamic DR creation. If INCLUDEIP and NODYNAMICDR are specified and data from the application does not include a DR request, the TIMEMARK method is used.

If chained data is sent from the host application, Telnet collects the entire chain before sending the complete data stream to the client. Telnet will save timestamp B when the entire data chain is sent.

It should be pointed out that the measured response time might not be exactly what the user sees. The first timestamp, A, is recorded when the data passes from Telnet to VTAM on its way to the SNA VTAM application. For various SNA reasons, Telnet might have received the data much earlier and had to queue it before it could be sent to VTAM. The most common reason for this is the application might not have given Telnet direction (the ability to send data). SNA applications often use a Change Direction Indicator (CDI) to manage which side can send data. The current sender can send data until it sends a CDI, which gives the other side permission to send. There might be scenarios where a client data request comes in, Telnet does not have direction, and must queue the data until the application sends a CDI. The measured response time will not include this queue time.

Average response time data collection

Specify AVERAGE within MonitorGroup to collect average response time data. Specify NOAVERAGE to turn off average response time data collection. If average response time data is requested, the following data is available on a per connection basis:

- Life-of-connection response time averages
- Sliding-window response time averages
- Variance and standard deviation of response time averages

Life-of-connection response time averages

These averages are based on data collected since the beginning of the connection. Data collected for this average includes:

- Total transaction count.
- Sum of the round-trip response times. Round-trip response time is the time difference between timestamp C and timestamp A.
- Sum of the IP response times. IP response time is the time difference between timestamp C and timestamp B. Sum of the SNA response times can be derived by subtracting the IP sum from the round-trip sum.

With this data collected, the average round-trip time, average IP time, and average SNA time can be calculated by dividing each sum by the transaction count. The data accumulation values might wrap. It is up to the management application to detect this. The Telnet connection display will indicate the wrapped condition instead of displaying averages.

Sliding-window response time averages

These averages give higher significance to more recent data without ever completely losing the impact of earlier data. The sliding-window methodology calculates an average over an interval of time instead of over the life of the connection. An interval is made up of a specified number of equal time periods. Use AVGSAMPMULTIPLIER to specify how many periods are in the interval. Use AVGSAMPPERIOD to specify the time length of a period. An interval should be long enough to have several data flows measured. Data collected during the period includes:

- Transaction count in the period (pTX).
- Sum of round-trip response times in the period (pRT).
- Sum of IP response times in the period (pIP). Sum of SNA response times can be derived.

In addition, there are three sliding-window variables that represent the interval and are used to calculate the sliding-window average. These variables are updated at the end of each period. They are:

- Sliding-window transaction count (swTX).
- Sliding-window sum of round-trip response times (swRT).
- Sliding-window sum of IP response times (swIP). Sliding-window sum of SNA response times can be derived.

When the interval has moved one period, subtracting the oldest period of collected data from the total and adding the newest period is one method for determining the new interval averages. However, this simple method loses all impact of earlier periods on the averages. Instead of dropping the collected data from the oldest period in the interval and completely losing its effect on the average, an average period of data is subtracted from the sliding-window totals. With this method, even the oldest period data continues to have a declining effect on the latest average calculations. And, by using an average period amount, the unique data for each period does not need to be retained. Only a total for the entire interval needs to be saved. Average period data is determined based on the fact that a period is a fraction of the total interval time. A period of data is 1/*n* of the current sliding-window totals, where *n* is the number of periods in the interval. The sliding-window average is derived from the new sliding-window values. At the end of each period the following calculations occur:

- Calculate the average period values:

```
avTX = swTX * (1/n)
avRT = swRT * (1/n)
avIP = swIP * (1/n)
```

- Remove an average period amount from the sliding-window totals and add the new period values:

```
swTX = swTX - avTX + pTX
swRT = swRT - avRT + pRT
swIP = swIP - avIP + pRT
```

- Calculate and report new sliding-window averages:

```
sliding-window round-trip average = swRT/swTX
sliding-window IP average = swIP/swTX
sliding-window SNA average = (swRT-swIP)/swTX
```

After the calculations are done, the period variables are reset and the next period of data collection begins. At the end of that period, the sliding-window variables are updated again with new information. The cycle continues for the life of the connection.

AVGSAMPMULTIPLIER 1 is a special case where 100% of the existing interval data is subtracted before the new period data is added. The result is an average for just that interval and the effect of older data is completely ignored.

AVGSAMPMULTIPLIER 0 is a special case that tells Telnet to include all data equally. The result is a life-of-connection average that is already available. A very large multiplier will have a similar effect.

A sliding-window round-trip average example follows. Assume an interval is made up of 5 periods. Each period is 2 minutes. Specify AVGSAMPMULTIPLIER 5 and AVGSAMPPERIOD 120. For illustrative purposes

in this example, the average response time starts low and then increases to a steady state value. The data collected for each period is the sum of all response times in milliseconds (ms) and the number of transactions (tx).

Table 29. Sliding-window round-trip average example

Period	Sum of response times (ms)	Number of transactions (tx)	Average response time (ms/tx)
1	1000	10	100
2	1650	15	110
3	2800	20	140
4	3500	25	140
5	4200	30	140
6	4900	35	140
7	5600	40	140

The sliding-window average response time for period 1:

```
1000ms/10tx = 1000ms/10tx = 100 ms/tx
```

The sliding-window average response time for period 2:

```
(1000-(1000*(1/5))+1650)ms/ (10-(10*(1/5))+15)tx = 2450ms/23tx = 106.5 ms/tx
```

The sliding-window average response time for period 3:

```
(2450-(2450*(1/5))+2800)ms/ (23-(23*(1/5))+20)tx = 4760ms/38.4tx = 124.0 ms/tx
```

The sliding-window average response time for period 4:

```
7308ms/55.7tx  = 131.2 ms/tx
```

The sliding-window average response time for period 5:

```
10046.4ms/74.6tx = 134.7 ms/tx
```

The sliding-window average response time for period 6:

```
12937.1ms/94.7tx = 136.6 ms/tx
```

The sliding-window average response time for period 7:

```
15949.7ms/115.7tx = 137.9 ms/tx
```

The sliding-window method continues to include the effects of the earlier response time averages. The greater the number of periods (AVGSAMPMULTIPLIER), the longer older data will continue to impact new average calculations. Decreasing the number of periods (AVGSAMPMULTIPLIER) gives less emphasis to older data. The AVGSAMPMULTIPLIER gives the system administrator control over the emphasis placed on old data.

Each of the last five periods in the example had average response times of 140 ms/tx. Different multiplier values give the following results after seven periods. It is clear that the greater the number of periods in the interval, the greater the impact of the older data.

- 2 period interval - 139.7 ms/tx
- 5 period interval - 137.9 ms/tx
- 10 period interval - 136.6 ms/tx

Variance and standard deviation of response time averages

It is useful to know the dispersion of the response time data. The variance is a measure of how variable the data is. The square root of the variance is the standard deviation. Assuming the distribution is a normal distribution, you can have the following confidence level:

- 68% that a response time value will fall within the average, plus or minus 1 standard deviation
- 95% that a response time value will fall within the average, plus or minus 2 standard deviations
- 99% that a response time value will fall within the average, plus or minus 3 standard deviations

The data that is saved for variance and standard deviation includes the data that is saved for life-of-connection averages, and the sum of each response time squared for the following data:

- Complete round-trip.
- IP portion of the transaction.
- SNA portion of the transaction. The sum of squares cannot be derived in this case and must be separately maintained.

These values might wrap. It is up to the management application to detect this. The Telnet connection display will indicate the wrapped condition instead of displaying these values or the standard deviation. The typical formula used for variance is:

```
VARIANCE = SUM((Xi-Xm)**2)/(n-1)
```

However, that would require keeping all the individual response times to calculate the sum of the squares. The same formula can be expanded and rewritten to not need individual response times. The formula used by Telnet is:

```
VARIANCE = [SUM(Xi**2) - (SUM(Xi))**2/n]/(n-1)
```

The values needed in this formula are saved by Telnet for each connection requesting average response time monitoring.

Time buckets

Five time buckets, each defined by a maximum response time, are used. Every transaction has a total response time, and that time fits into one of the five configurable time buckets. The bucket boundaries are created by specifying the maximum response time for the first four buckets. The fifth bucket has an open-ended maximum value. The minimum response time boundary for each bucket is the maximum of the previous bucket. The first bucket has a minimum value of 0. A bucket transaction count is incremented by one when a response time is greater than the minimum and less than or equal to the maximum. Figure 93 on page 654 depicts the time buckets. For default values, see z/OS Communications Server: IP Configuration Reference.

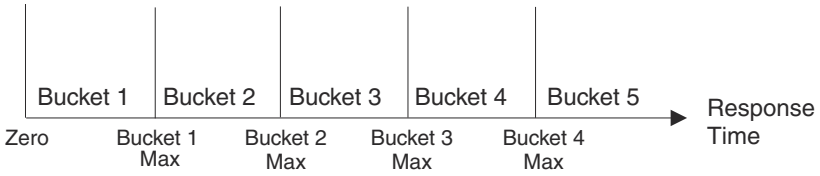


Figure 93. Time buckets

Reducing demand for ECSA storage

Telnet server configurations that support a large number of connections can place a high demand on ECSA storage. To reduce this demand, configure the Telnet server to enable multiple Telnet LUs to share an ACB.

To configure the Telnet server to enable LUs to share an ACB, specify SHAREACB in the TELNETGLOBALS statement block. Replace any predefined (static) VTAM APPL definitions that are being used to represent

Telnet LUs with corresponding model application program definitions; the latter are required when using the SHAREACB function.

Configuring the z/OS UNIX Telnet server

The z/OS UNIX Telnet server (otelnetd) provides access to z/OS UNIX shell applications on the host using the Telnet protocol. The z/OS UNIX Telnet server lets hosts in an IP network log on to the z/OS shell environment directly, without going through TSO. The z/OS UNIX Telnet server supports AIX and UNIX full-screen applications such as the vi editor, so that AIX and UNIX users can use familiar Telnet commands. The z/OS UNIX Telnet server runs in both line mode and raw mode, but does not support TN3270 or TN3270E, as the TN3270E Telnet server does.

Installation information

The files used in the z/OS UNIX Telnet server and their locations in the z/OS UNIX file system are as follows:

/etc/services
The ports for each application are defined here. For example:

```
otelnet xxxx/tcp
```

where *xxxx* is the port that inetd should listen on for otelnet.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file. otelnetd writes to syslog facility local1.

/etc/inetd.conf
The configuration parameters for all applications started by inetd are defined in this file.

/usr/sbin/otelnetd
This is a symbolic link to /usr/lpp/tcpip/sbin/otelnetd. /usr/lpp/tcpip/sbin/otelnetd is a sticky-bit file. The OTELNETD member of SEZALOAD contains the executable code for the Telnet server.

/etc/banner
This file contains a login message that is printed to the client's screen after the client logs in, unless the -h option is specified. Store the banner in this file.

/etc/otelnetd.banner
This file contains a message that is printed to the client's screen prior to the login prompt when the user connects to the server, unless the -h option is specified. Store the banner in this file.

/etc/utmpx
This file is updated by the call to fsumoclp. It contains a list of all the users who are logged in with their associated tty.

/dev/ptypXXXX and /dev/ttypXXXX
These special device files represent pseudoterminals (ptys); they are used by the z/OS UNIX Telnet server and other programs.

Note: For information on allocating more of these files for more connections, see z/OS UNIX System Services Planning.

/usr/share/lib/terminfo
The descriptions of supported terminals are stored here. For more information, see z/OS UNIX System Services Planning.

/usr/lib/nls/msg/C/tnmsgs.cat
The message catalog used by the z/OS UNIX Telnet server is stored here.

If the message catalog does not exist, the software will use the messages hard-coded within the software by default. These messages duplicate the English message catalog that is shipped with the product.

/usr/man/C/cat1/otelnetd.1
This file contains the associated manual (man) pages for the z/OS UNIX Telnet server. It provides online help for the user.

Environment variables

Table 30 on page 656 provides a list of environment variables that can be explicitly set by z/OS UNIX Telnet.

Table 30. Environment variables for z/OS UNIX Telnet

Environment variable	Description
_BPX_SHAREAS	Controls whether a spawned child process is started in the same address space as the login shell.
KRB5_SERVER_KEYTAB	Specifies the location of the key table.
LC_ALL	Determines the values for all local categories.
LC_COLLATE	Determines the local category for character collation.
LC_CTYPE	Determines the local category for character handling functions, such as tolower(), toupper(), and isalpha(). Also determines the interpretation of sequences of bytes of text data as characters (for example, single as opposed to multibyte characters), the classification of characters (for example, alpha, digit, graph), and the behavior of character classes.
LC_MESSAGES	Determines the local category for processing affirmative and negative responses, and the language and cultural conventions in which messages should be written.
LC_NUMERIC	Determines the local category for numeric formatting information (for example, thousands separator and radix character) in various utilities, as well as the formatted I/O operations in printf() and scanf() and the string conversion functions in strtod().
LC_TIME	Determines the local category for date and time formatting information. It affects the behavior of the time functions in strftime(). Additional semantics of this variable, if any, are implementation defined.
NLSPATH	Contains a sequence of templates that the catopen() function uses when attempting to locate message catalogs. Each template consists of an optional prefix, one or more conversion specifications, a file name, and an optional suffix.
TERMINFO	Specifies the path name for an unsupported terminal that has been added to the terminfo file. Use the TERMINFO variable in /etc/profile or /etc/.login.

Starting, stopping, and administration of z/OS UNIX Telnet

The z/OS UNIX Telnet server is started by inetd for each incoming Telnet connection. When the Telnet session is complete, the z/OS UNIX Telnet server will exit. Each active Telnet session will have a separate instance of the Telnet server which will communicate with the Telnet client.

The z/OS UNIX inetd daemon does not propagate environment variables other than PATH and TZ to its child processes, so the NLSPATH and LANG environment variables cannot be used to point to a different message catalog.

The following standards are supported:

- RFC 854 Telnet Protocol Specification
- RFC 855 Telnet Option Specification

- RFC 856 Telnet Binary Transmission
- RFC 857 Telnet Echo Option
- RFC 858 Telnet Suppress Go Ahead Option
- RFC 859 Telnet Status Option
- RFC 860 Telnet Timing Mark Option
- RFC 861 Telnet Extended Options - List Option
- RFC 885 Telnet End of Record Option
- RFC 1073 Telnet Window Size Option
- RFC 1079 Telnet Terminal Speed Option
- RFC 1091 Telnet Terminal type option
- RFC 1096 Telnet X Display Location Option
- RFC 1123 Requirements for Internet Hosts -- Application and Support
- RFC 1184 Telnet Linemode Option
- RFC 1372 Telnet Remote Flow Control Option
- RFC 1571 Telnet Environment Option Interoperability Issues
- RFC 1572 Telnet Environment Option
- RFC 2941 Telnet Authentication Option
- RFC 2942 Telnet Authentication: Kerberos Version 5
- RFC 2946 Telnet Data Encryption Option
- RFC 2952 Telnet Encryption: DES 64 bit Cipher Feedback
- RFC 2953 Telnet Encryption: DES 64 bit Output Feedback

When a z/OS UNIX Telnet session is started up, otelnetd sends Telnet options to the client side that indicate a willingness to do the following options:

- WILL ENCRYPT
- DO ENCRYPT
- DO TERMINAL TYPE
- DO TSPEED
- DO XDISPLOC
- DO NEW-ENVIRON
- DO ENVIRON
- WILL SUPPRESS GO AHEAD
- DO ECHO
- DO LINEMODE
- DO NAWS
- WILL STATUS
- DO LFLOW
- DO TIMING-MARK

The z/OS UNIX Telnet server can enable the following options locally.

- WILL BINARY

  This option indicates that the client is willing to send 8 bits of data, rather than the normal 7 bits of network virtual terminal data.
- WILL ECHO

When the LINEMODE option is enabled, a WILL ECHO or WONT ECHO is sent to the client to indicate the current state of terminal echoing. When terminal echo is not wanted, a WILL ECHO is sent to indicate that Telnet takes care of echoing any data that must be echoed to the terminal, and then nothing is echoed. When terminal echo is wanted, a WONT ECHO is sent to indicate that Telnet is not doing any terminal echoing, so the client can do any terminal echoing that is needed.

- WILL LOGOUT

  When a DO LOGOUT is received, a WILL LOGOUT is sent in response and the Telnet session is shut down.

- WILL SGA

  This option indicates that it will not be sending IAC GA, the go ahead command.

- WILL STATUS

  Indicates a willingness to send the client, upon request, the current status of all Telnet options.

- WILL TIMING-MARK

  Whenever a DO TIMING-MARK is received, a WILL TIMING-MARK is the response. It is used only in kludge linemode support.

- WILL ENCRYPT

  Indicates a willingness to encrypt the data stream.

The z/OS UNIX Telnet server can enable the following options remotely.

- DO BINARY

  Sent to indicate that Telnet is willing to receive an 8-bit data stream.

- DO ECHO

  If a WILL ECHO is received, a DONT ECHO will be sent in response.

- DO ENVIRON

  Indicates a desire to be able to request environment variable information. (See RFC 1408.)

- DO LFLOW

  Requests that the client handle flow control characters remotely.

- DO LINEMODE

  Supports requests that the client do line-by-line processing.

- DO NAWS

  Requests that the client inform the server when the window size changes.

- DO NEW-ENVIRON

  Indicates a desire to be able to request environment variable information. (See RFC 1572.)

- DO SGA

  Indicates that it does not need to receive IAC GA, the go ahead command.

- DO TERMINAL-TYPE

  Indicates a desire to be able to request the name of the type of terminal that is attached to the client side of the connection.

- DO TERMINAL-SPEED

  Indicates a desire to be able to request information about the speed of the serial line to which the client is attached.

- DO TIMING-MARK

  Only supported if the client responded with WONT LINEMODE. If the client responds with WILL TM, then it is assumed that the client will support kludge linemode. It is not used for any other purposes.

- DO XDISPLOC

  Indicates a desire to be able to request the name of the X Window System display that is associated with the Telnet client.
- DO AUTHENTICATION

  Indicates a willingness to receive authentication information for automatic login.
- DO ENCRYPT

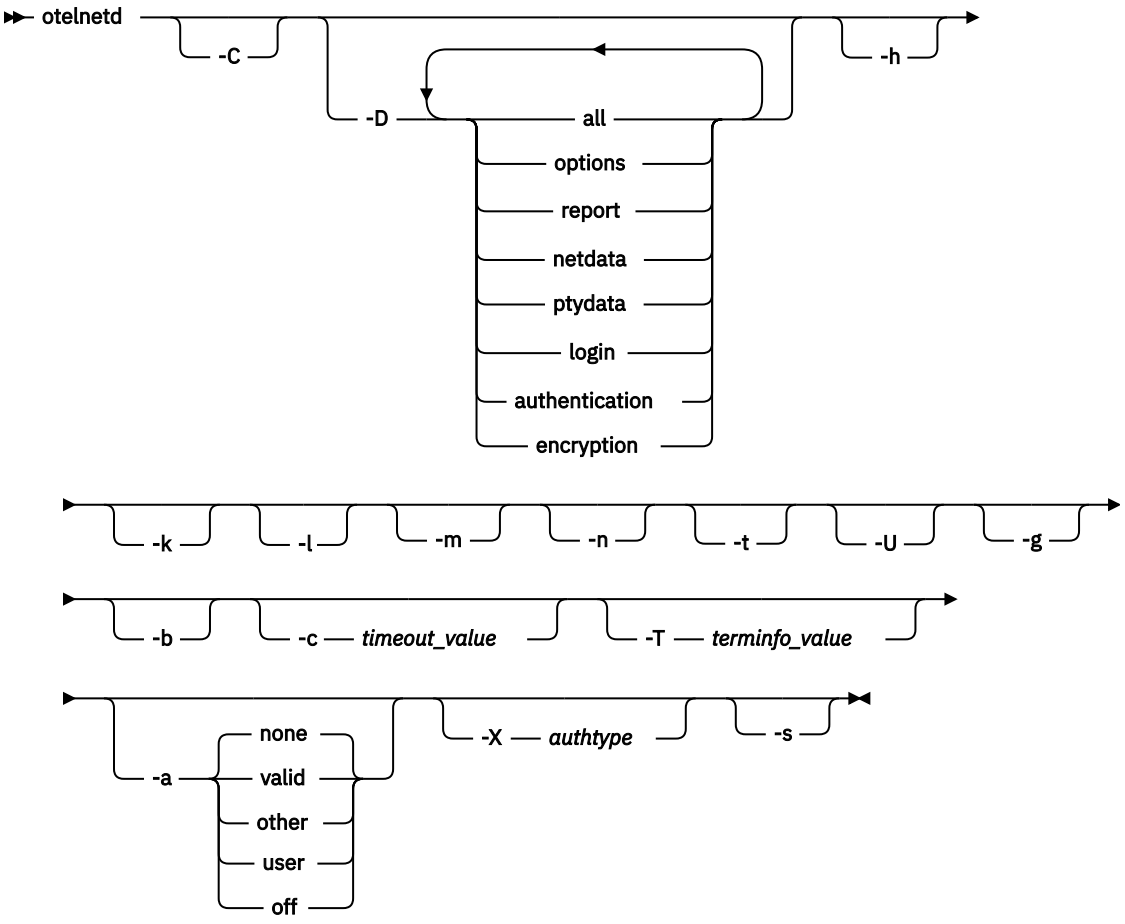
  Indicates a willingness to decrypt the data stream.

otelnetd

Note: The user ID associated with the daemon in /etc/inetd.conf requires superuser authority. See z/OS UNIX System Services Planning for a description of the types of authority defined for daemons.

The following syntax is used in the /etc/inetd.conf file to define the arguments used to invoke otelnetd.

Syntax



Parameters

-C

Prints user messages in uppercase. There are several exceptions. Messages issued at startup are not affected by the -C option because the -C option is not processed during the startup. Also, data transmittal messages will not be uppercase. Data transmittal messages are generated from the -D netdata option or the -D ptydata option.

-D

The following suboptions apply to -D:

options
Prints information about the negotiation of Telnet options. This information is used for debugging purposes. This suboption allows telnetd to generate debugging information to the connection, which allows the user to view telnetd activity.

report
Prints the options information and additional information about processing. This information also includes print information designated for suboption=options. This can be used for debugging purposes. This suboption telnetd to generate debugging information to the connection, which enables the user to view telnetd activity.

netdata
Displays the data stream received by telnetd. This information is used for debugging purposes. It allows telnetd to generate debugging information to the connection, which enables the user to view telnetd activity.

ptydata
Displays the data stream written to the pty. This information is used for debugging purposes. It allows telnetd to generate debugging information to the connection, which enables the user to view telnetd activity.

all
Enables options, report, netdata, ptydata, login, authentication and encryption.

login
Records login and logout activity to syslogd facility auth using message EZYTU36I.

authentication
Turns on authentication debugging code.

encryption
Turns on encryption debugging code.

-h

Disables the display of the /etc/banner and /etc/otelnetd.banner files at the terminal of the client.

-k

Disables kludge linemode. The server normally attempts to use kludge linemode when the `-l` option was specified, but the client does not support line mode. Use the `-k` option when there are remote clients that do not support kludge linemode, but pass the heuristic for kludge line mode support (for example, if they respond with WILL TIMING-MARK in response to a DO TIMING-MARK). This option does not disable kludge line mode when the client requests it. This is accomplished by the client sending DONT SUPPRESS-GO-AHEAD and DONT ECHO.

-l

Specifies linemode, which tries to force clients to use linemode. If the LINEMODE option is not supported and the `-k` option was not specified, it will attempt to use kludge linemode.

Notes:

1. Many clients decline the server's request to operate in linemode.
2. Linemode is not appropriate for full-screen applications like the z/OS UNIX vi editor.

-m

Enables the creation of a forked or spawned process to coexist in the same address space. This option can improve performance because the user's login shell runs in the same address space as otelnetd.

-n

Disables TCP keep-alives. Normally, telnetd enables the TCP keep-alive mechanism to probe connections that have been idle for some time to determine if the client is still there. In this way, idle connections from machines that have crashed or can no longer be reached can be cleaned up. The cleanup of disabled connections is controlled by the presence of the INTERVAL parameter on the TCPCONFIG statement in the TCPIP profile.

-t
Specifies internal tracing. It also activates the REPORT option, as if the user also specified -D Report.

-U
Causes telnetd to drop connections from any IP address that cannot be mapped back into a symbolic name by the gethostbyaddr or getnameinfo routines.

Result: If coded, the -U parameter causes the -g parameter to be ignored.

-g
Disables the ability to issue the gethostbyaddr or getnameinfo routines that use the client IP address to resolve the client host name.

Results:

- If this parameter is coded, the host name does not appear in the trace output (-t parameter) or in the WHO command output.
- This parameter is ignored if the -U parameter is coded.

-b
Forces the server to DO BINARY in the first pass during negotiations with the client.

-c *timeout_value*
Specifies the number of seconds to wait before terminating the Telnet session for inactive connections. The *timeout_value* is a value between 1 and 86400 seconds.

-T *terminfo_value*
Sets the TERMINFO environment variable to the specified values at startup. This option is needed when terminfo definitions are located in nonstandard directories.

-a
This option may be used for specifying what mode should be used for authentication. There are several valid suboptions for authentication mode:

valid
Allow connections only when the remote user can provide valid authentication information to identify the remote user. Thus, for otelnetd, Kerberos authentication will be required. User verification will still occur through the login and password prompt. However, if the login user ID matches the TSO user ID that was mapped from the name in the Kerberos principal using the SAF R_usermap function, then no password will be requested. This is the most secure authentication mode.

other
Allow only connections that supply some authentication information. This option is currently not supported by any of the existing authentication mechanisms, and is thus the same as specifying `-a valid`.

user
Allow connections only when the remote user can provide valid authentication information to identify the remote user, and is allowed access to the specified account without providing a password. Thus, for otelnetd, Kerberos authentication is required. The NAME received during AUTHENTICATION option negotiation must match the name in the Kerberos principal, and the Kerberos principal must map to a valid TSO user ID on the host using the SAF R_usermap function. No user verification will occur through the login or password prompt.

none
This is the default state. Authentication information is not required. User verification will still occur through the login and password prompt. However, if the login user ID matches the TSO user ID that was mapped from the name in the Kerberos principal using the SAF R_usermap function, then no password will be requested.

off
This disables the authentication code. All user verification happens through the login and password prompt. During option negotiation, otelnetd will not send DO AUTHENTICATION and, if necessary, will send DONT AUTHENTICATION.

Note: Authentication is not supported for IPv6 connections. If tcp6 is specified in inetd.conf, -a should not be used as a start option. If tcp6 and -a are both specified, the suboption will be overridden and forced to OFF.

-X *authtype*
This option disables the use of authtype authentication. Currently the only valid value for authtype is KERBEROS_V5. Thus, if otelnetd sends the AUTHENTICATION option SEND command, the authentication-type-pair-list will not contain any KERBEROS_V5 entries and will be empty.

-s
Used to set the KRB5_SERVER_KEYTAB environment variable. If this environment variable is set, run time security uses a local instance of the Kerberos security server to decrypt service tickets instead of obtaining the key from a key table. To use this capability, the otelnetd application must have at least READ access to the IRR.RUSERMAP resource in the FACILITY class. For more information, see z/OS Integrated Security Services Network Authentication Service Administration.

SMF record handling

The SMF records generated are the typical set of records that MVS generates for start of job (login) and end of job (logoff). Additionally, interval records might be issued during the life of the user login. These records are SMF type 30 and type 72 and not the type 118 or type 119 in the current z/OS UNIX Telnet server. The process of issuing these records is external to the specific daemons.

BPX.DAEMON considerations

If the BPX.DAEMON FACILITY class profile is defined, perform the following additional configuration steps:

1. Provide read access to BPX.DAEMON for the user ID specified in /etc/inetd.conf for otelnetd.
2. Define SEZALOAD to program control.
3. Define the C run-time library, *hlq*.SCEERUN to program control.

See z/OS UNIX System Services Planning for more information about the BPX.DAEMON FACILITY class profile, the security product commands used to perform the required configuration, and the diagnosis procedure for resolving related problems.

Kerberos

otelnetd supports Kerberos Version 5 for authentication on IPv4 connections. Authentication is not supported on IPv6 connections (that is, if tcp6 is specified for otelnetd in inetd.conf). On z/OS, Kerberos is implemented by Security Server. See z/OS Integrated Security Services Network Authentication Service Administration for more information.

If you plan to use Kerberos authentication with otelnetd, you need to ensure that the user ID under which inetd runs has permission to the ICSF callable services that Kerberos needs. Specifically, if the CSFSERV class is active, ensure that the user ID of inetd has read access to the following ICSF SAF resources:

CSF1TRC, CSF1SKE, CSF1SKD, CSF1TRD, and CSFOWH

The Kerberos principal used by otelnetd will generally be of the form "host/<*hostname*>@*realm*". That is, the first component of the Kerberos principal is "host"; the second component is the fully qualified lowercase hostname of the server; and the realm is the Kerberos realm to which the server belongs.

otelnetd will not accept forwarded credentials from the client.

Successful AUTHENTICATION option negotiation is required for successful ENCRYPT option negotiation. The ENCRYPT option must be negotiated in both directions.

Chapter 12. Transferring files using FTP

The File Transfer Protocol (FTP) allows a user to copy files from one machine to another. The protocol allows for data transfer between the client (the user) and the server in either direction. In addition to copying files, the client can issue FTP commands to the server to manipulate the underlying file system of the server (for example, to create or delete directories, delete files, rename existing files, and so on.) FTP is the most common TCP/IP application for moving files between computers.

Copying files from one machine to another is one of the most frequently used operations. The data transfer between client and server can be in either direction. The client can send a file to the server machine. It can also request a file from this server.

To access remote files, the user must identify himself or herself to the server. At this point the server is responsible for authenticating the client before it allows the file transfer.

From an FTP user's point of view, the link is connection-oriented. FTP uses TCP as a transport protocol to provide reliable end-to-end connections. Both hosts must run TCP/IP to establish file transfer.

The z/OS model for the FTP server includes a daemon process and a server process. The daemon process starts when you start your cataloged procedure (for example, START FTPD) and it listens for connection requests on a specific port. The port is the well-known port 21 unless otherwise specified. For methods of choosing a different port number, see “Configuring ETC.SERVICES” on page 665 and “Configuring the FTPD cataloged procedure” on page 665. When the daemon accepts an incoming connection, it creates a new process (server's address space) for the FTP server, which handles the connection for the rest of the FTP login session. Each login session has its own server process.

The server process inherits the accepted connection from the daemon process. This connection is called the control connection. The server receives commands from the client and sends replies to the client using the control connection. The control connection port is the same as the daemon's listening port.

The client and server use a different connection for transferring data; this connection is called the data connection. By default, the data port is one less than the control connection port. For example, if the control connection port is 21, the data port is 20. An FTP client can override the default data port by directing the server to run in passive mode. In passive mode, the server uses an ephemeral port for the data port. Passive mode is requested by firewall friendly clients and by clients initiating three-way data transfers.

Notes:

1. This topic discusses RACF configuration required for FTP. References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on configuration.
2. If you use the environment variable _BPX_JOBNAME when you start FTPD, the server's address space is known as the job name specified in the _BPX_JOBNAME variable. You might need to have a common naming convention for your installation's FTP address spaces if your installation uses syslogd isolation or has other workload management requirements.

   If you do not use the _BPX_JOBNAME environment variable, the server's address space assumes the name of the user. For example, if a user logs into an FTP server with the user ID TCP0001, the FTP server address space servicing the request is also known as TCP0001.

   If the FTP daemon accepts a connection that is protected by the TLS security mechanism and you are not using the _BPX_JOBNAME environment variable, the server's address space name is a name derived from the FTP server job name. The name is in the form *jobnamex*, where the *jobname* value is the job name, and the *x* value is a number in the range 1 – 9. If the FTP daemon accepts a connection that is protected by the TLS security mechanism and you are using the _BPX_JOBNAME environment variable, the server's address space name is a name derived from the _BPX_JOBNAME environment variable. The name is in the form *bpxjobnamex*, where the *bpxjobname* value is the value specified for the _BPX_JOBNAME environment variable, and the *x* value is a number in the range 1-9.

© Copyright IBM Corp. 2000, 2022

Configuring PROFILE.TCPIP for FTP

If you have configured the FTP server to have affinity to a specific stack, or you have configured the FTP server to be a generic server in a single stack environment, the FTP server can be started automatically when the TCP/IP address space is started by specifying the name of the FTP server cataloged procedure in the AUTOLOG statement. If you have configured the FTP server as a generic server in a multiple stack environment, you should not use the AUTOLOG statement to automatically start the server. Instead, use some other automation outside of AUTOLOG to automatically start the server.

In the following example, if your procedure is called FTPD, the following statement allows TCP/IP to issue the MVS START command for procedure FTPD. The job name of FTPD1 will be used on the port statement shown below. If the daemon job name is fewer than eight characters, the FTP daemon forks a process that has the job name of the original daemon appended with the numeral 1.

```
AUTOLOG
  FTPD JOBNAME FTPD1
ENDAUTOLOG
```

To reserve ports 21 and 20 for the FTP server, add the following statement:

```
 PORT
   21 TCP FTPD1             ; FTP server control port
   20 TCP OMVS NOAUTOLOG    ; FTP server data port
```

Specifying FTPD1 on the PORT and AUTOLOG statements directs TCP/IP to restart FTPD if it shound end.

To allow FTP to detect data connection errors when there has been no activity on the data connection for a certain amount of time, set the INTERVAL parameter on the TCPCONFIG statement to a relatively low value. The keepalive packets that the stack sends as specified on the INTERVAL parameter enable the stack to detect errors, such as a reset or terminated peer connection, instead of waiting indefinitely. Be careful when choosing an INTERVAL value on the TCPCONFIG statement because this value will affect all TCP connections at the host for which the interval has been activated, not just FTP connections.

The control connection can also benefit from keepalive packets. Many firewalls require periodic activity on any connection that is made and the control connection can appear idle during a long data transfer. Coding the INTERVAL parameter on the TCPCONFIG statement will, of course, cause keepalive packets to be sent on the control connection as well as the data connection. You can override the keepalive interval that you have configured in the stack for the FTP control connection and data connection with the FTPKEEPALIVE (control connection) and DATAKEEPALIVE (data connection) statements in the FTP.DATA file or data set.

Optimally, FTP needs a buffer size of at least 180K for data connections. You should not set the TCPMAXRCVBUFRSIZE parameter below 180K. The default value for the parameter is 256K. IBM Health Checker for z/OS can be used to check whether the TCPMAXRCVBUFRSIZE value is sufficient to provide optimal support to the z/OS Communications Server FTP server. For more details about IBM Health Checker for z/OS, see z/OS Communications Server: IP Diagnosis Guide and IBM Health Checker for z/OS: User's Guide.

For more information about the AUTOLOG, PORT, and TCPCONFIG statements, see z/OS Communications Server: IP Configuration Reference.

If your FTP server accepts connections on a distributed Dynamic VIPA (DVIPA), SYSPLEXPORTS must be specified for the distributed DVIPA if either of the following conditions are true:

- The DVIPA is an IPv6 address.
- Passive mode FTP is used.

For more information about the VIPADYNAMIC statement and specifying the SYSPLEXPORTS option on the VIPADISTRIBUTE parameter, see z/OS Communications Server: IP Configuration Reference.

Configuring ETC.SERVICES

The ETC.SERVICES file contains the relationship between service names (servers) and port numbers in the z/OS UNIX environment. If necessary, update your ETC.SERVICES file to include the control port that the FTP server is to use. For the search order used to locate the ETC.SERVICES file, see "Configuration files for TCP/IP applications" on page 25. For example, add the following line:

```
ftp 21/tcp
```

Notes:

1. In the ETC.SERVICES file, only one port (the one for the control connection) is listed.
2. If the ETC.SERVICES file is changed such that a port other than 21 is specified, that port will become the FTP port for that z/OS host.
3. The port specified for FTP in the ETC.SERVICES file can be overridden by the FTP start parameter, PORT *nnnn*. In either case, the port that is specified should match the port specified for FTP on the PORT statement in PROFILE.TCPIP.

Configuring /etc/syslog.conf

Note: For FTP syslog, you should consider the fact that FTP writes log messages to the system console if syslogd is not running. If you enable FTP server traces without syslogd active, large amounts of data might be written to the system console.

The daemon.*priority* entries in /etc/syslog.conf determine where FTP messages and trace entries are written. The FTP server issues info, warning, and error messages. All trace entries are written with debug priority. To direct trace entries (and all messages) to /tmp/daemon.trace, include the following line in /etc/syslog.conf:

```
*.*.daemon.debug    /tmp/daemon.trace
```

Log messages can be isolated within syslogd. For FTP, an installation might want FTP log messages to be written to different files depending on the user ID, or separately for the FTP daemon. If FTP messages are to be isolated for user1, use the first statement below. If FTP messages are to be logged for all the FTP applications, use the second statement below.

```
user1.*.daemon.debug /tmp/daemon.trace
```

```
*.FTPD*.daemon.debug /tmp/daemon.trace
```

In this statement, it is assumed that _BPX_JOBNAME is set to FTPD.

Configuring the FTPD cataloged procedure

The FTPD catalogued procedure is sample JCL that you can use to start the FTP server. To use the sample, you must modify it to suit your needs.

Before you begin

You must configure TCP/IP and your security product. You will modify the configuration of your security product for use with your FTP server.

Procedure

Perform the following steps to configure the FTPD catalogued procedure:

1. Copy the sample in SEZAINST(FTPD) to your system or recognized PROCLIB.
2. Update the SYSFTPD DD and SYSTCPD statements.

See “Configuring FTP.DATA” on page 676 to configure SYSFTPD DD and “Configuring TCPIP.DATA for FTP ” on page 675 to configure SYSTCPD DD.

3. Decide whether you will pass start parameters in the FTPD catalogued procedure.
   - If you are not going to pass start parameters, add your parameters to the FTP.DATA file.
   - If you are going to pass start parameters, add your parameters to the PARMS parameter in the PROC statement of the FTPD cataloged procedure.

     Any parameters that you modify in the PROC statement override parameters that are set in the FTP.DATA file.

     The system parameters required by the FTP server are passed by the PARM parameter on the EXEC statement of the FTPD cataloged procedure. For example, the entry `//FTPD PROC MODULE='FTPD',PARMS='TRACE ANONYMOUS PORT 21'` starts FTP with tracing active, anonymous support enabled, and using control port 21.

   For more information about the FTP server cataloged procedure (FTPD) parameters, see z/OS Communications Server: IP Configuration Reference.
4. Define the FTPD catalogued procedure to the security program.

   Add the FTPD catalogued procedure to the RACF STARTED class facility or to the started procedures table.

   The user ID that is associated with the FTP server STARTED class must have UID 0. If the FACILITY class is active and the BPX.DAEMON or BPX.POE profiles are defined, the user ID that is associated with the FTP server must have READ access to them.
5. (Optional) If the daemon address space will be configured to run as nonswappable, provide at least READ access to the FACILITY class resource BPX.STOR.SWAP.
6. Set up security for the FTP server.

   See “Security for the FTP server” on page 666 for information about setting up security.
7. (Optional) Define environment variables.

   See “Defining environment variables for the FTP server (optional)” on page 673 for information about defining environment variables for the FTP server.

Results

When you are done, you should be able to start the FTPD catalogued procedure with no errors. If you receive errors, then ensure that you have completed all the steps correctly.

See SEZAINST(EZARACF) for more information about SAF resource requirements needed for FTP.

Restriction: The Language Environment runtime option NATLANG(JPN) is not supported. If you specified NATLANG(JPN) as a Language Environment runtime option, then you need to specify PARM='NATLANG(ENU)' in the FTPD cataloged procedure to override the runtime option for FTP.

Security for the FTP server

To provide security for the FTP server, you must perform the following tasks:

1. (Optional) Activate and define the SERVAUTH class (see “(Optional) Steps for activating and defining the SERVAUTH class” on page 667).
2. Set up security for the FTP server (see “Steps for setting up security for your FTP server” on page 667).
3. Provide and control user access to the FTP server (see “Steps for controlling user access to the FTP server” on page 668).
4. Set up a port of entry for users of the FTP server (see “Steps for setting up a port of entry for users of the FTP server” on page 669).
5. Provide and control user access to the z/OS UNIX file system (see “(Optional) Steps for controlling user access to the z/OS UNIX file system” on page 670).

6. Prevent exploitation of your FTP server (see “Preventing exploitation of your FTP server” on page 671).
7. (Optional) Assign password phrases to user IDs that are used to log in to the FTP server (see “(Optional) Assigning password phrases to user IDs that are used to log in to the FTP server” on page 672).

FTP uses resource profiles in the System Authorization Facility (SAF) SERVAUTH class to control access to certain facilities and servers. When access to a resource is controlled by a profile in the SERVAUTH class, you must activate and RACLIST the SERVAUTH class. You do not have to use the SERVAUTH class, but when a profile is defined in that class, all FTP users who require access to it must be permitted to it.

For more information, see z/OS UNIX System Services Planning and z/OS Security Server RACF Security Administrator's Guide. For more information about network access security zones, see “Network access control” on page 140. If you are planning to implement a multilevel security environment on your z/OS system, see Chapter 4, “Preparing for IP networking in a multilevel secure environment,” on page 191.

(Optional) Steps for activating and defining the SERVAUTH class

FTP uses resource profiles in the System Authorization Facility (SAF) SERVAUTH class to control access to certain facilities and servers.

Before you begin

You need to know which resource profiles you want to define. You need to install and start your security product.

Procedure

Perform the following steps to activate and RACLIST the SERVAUTH class, if you have not already done so:

1. Issue the following command from a RACF special user to activate the SERVAUTH class:

   ```
   SETROPTS CLASSACT (SERVAUTH)
   ```

   Requirement: If you change the SERVAUTH class after you activate it, you must refresh the class. Changes include, but are not limited to, adding a resource profile to the SERVAUTH class or changing access to a profile in the SERVAUTH class. To refresh the class, issue the following command from a RACF special user:

   ```
   SETROPTS RACLIST (SERVAUTH) REFRESH
   ```

2. Issue the following command from a RACF special user to RACLIST the SERVAUTH class:

   ```
   SETROPTS RACLIST (SERVAUTH)
   ```

Steps for setting up security for your FTP server

You need to provide appropriate access to the user ID that is associated with the FTP daemon.

Before you begin

You need to know the user ID that is associated with the FTP daemon and how TCP/IP is configured for security. You should also know the resource profiles that are in the SAF classes.

Procedure

To set up security for your FTP server, perform one or more of the following tasks:

- If the SERVAUTH class is activated and a profile is defined for the EZB.STACKACCESS.*mvsname.tcpname* resource, you must grant the user ID that is associated with the FTP daemon READ access to the profile.

- If the SAF class APPL is activated and the OMVSAPPL resource profile is defined, grant the user ID that is associated with the FTP daemon READ access to the OMVSAPPL resource profile.

  For more information on defining the OMVSAPPL profile, see z/OS UNIX System Services Planning.
- If the SAF class APPL is activated and you have a resource profile defined in that class that matches the job name of the address space that the FTP server starts when a user logs into FTP, a user ID should have READ access to that resource profile.
- The FTP daemon listening port should be reserved for the FTPD job by a PORT statement in the TCPIP PROFILE. If the PORT statement for the FTPD port is protected with the SAF keyword, you must define a SERVAUTH profile for the EZB.PORTACCESS.*sysname*.*tcpname*.SAFkeyword resource. The user ID associated with the FTP daemon must have READ access to that resource.
- If your IP network is configured to use named security zones, grant the user ID that is associated with the FTP daemon READ access to the security zone that maps its bind address (0.0.0.0/32 for INADDR_ANY or ::/128 for the IPv6 unspecified address, in6addr_any), unless these addresses are overridden by the PORT statement in the TCP/IP profile.

Results

You know you are done when you can start the FTP server without receiving an error.

Steps for controlling user access to the FTP server

Every user that logs in to your FTP server requires access to that server. Use these steps to provide and control user access to your server.

Before you begin

You need to know which users you want to allow to log in to your FTP server. You need to know whether your IP network is configured to use named security zones.

Procedure

Perform the following steps to control user access to the FTP server:

1. Provide each user who is going to log in to the FTP server with a z/OS UNIX UID.

   You can either provide a UID to the user, or the user can use the default UNIX UID.
2. If your IP network is configured to use named security zones, each defined security zone has a SERVAUTH profile for the resource named EZB.NETACCESS.*sysname*.*tcpname*.*zonename*. If the client IP address is mapped into a network access security zone, grant each login user ID READ access to the SERVAUTH profile that corresponds to the security zone.

   For more information about security zones, see “Network access control” on page 140.
3. Do one or more of the following items to allow only certain users to log in to the FTP server:
   - Code an FTCHKPWD user exit routine to allow or deny access to users, based on user ID.

     For more information about user exits, see FTP server user exits in z/OS Communications Server: IP Configuration Reference and “Configuring the optional FTP user exits” on page 708.
   - Use the SERVAUTH resource profile that FTP uses for TLS level 3 authentication to control which users can log in to FTP:

     a. Define a profile in the SERVAUTH class for the FTP port.

     b. Grant at least READ access to the profile to the users that you want to permit to log in to FTP.

        For example, if your security product is RACF, your FTP port is port 21, and the profile that you defined is EZB.FTP.*.*.PORT21, issue the following command to grant the user ID FTPUSER access to the profile:

        ```
        PERMIT EZB.FTP.*.*.PORT21 CL(SERVAUTH) ID(FTPUSER)
        ```

See z/OS Security Server RACF Command Language Reference, z/OS Security Server RACF Security Administrator's Guide, or the documentation for your SAF-compliant security product for more information.

c. Code VERIFYUSER TRUE in the server's FTP.DATA file.

FTP verifies the user's access to the profile for every session, whether or not that session is secured. TLS-secured sessions are also verified, even when level 3 authentification has not been requested.

4. (Optional) Set up transport layer security (TLS) support or Kerberos support for the FTP server.

The FTP server supports TLS. TLS enables secure file transfer by providing data privacy, message authentication, and message integrity services for data sent and received using the FTP control and data connections. For information about setting up TLS support for the FTP server, see "Customizing Transport Layer Security and Kerberos security" on page 686.

You can use the Generic Security Service Application Programming Interface (GSSAPI) to authenticate FTP clients to FTP servers. For more information about setting up GSS support for the FTP server, see "Customizing Transport Layer Security and Kerberos security" on page 686.

Results

When you are finished, only certain users will be able to log in to your FTP server.

Steps for setting up a port of entry for users of the FTP server

The *port of entry* is the origin of work for the FTP server. You must establish a port of entry for each user who logs in to your FTP server.

Before you begin

You must know:

- The IP addresses of the clients that are to log in to your FTP server
- Whether your connection partners are in a network access security zone
- Whether your RACF SETROPTS options are TERMINAL(READ) or TERMINAL(NONE)

For IPv4 connection partners, you can establish either terminal access or SERVAUTH access. IPv6 connection partners must use SERVAUTH access, which is established automatically for them.

Procedure

Perform the following steps to set up the port of entry for IPv4 and IPv6 users of the FTP server.

- To establish terminal access for IPv4 connection partners, take one of the following actions:
  - If your RACF SETROPTS options are TERMINAL(NONE):
    1. Define profiles for the IP addresses that you want to permit to your system in the TERMINAL class.

       Translate all the IP addresses of any clients that connect to the FTP server to an 8-byte hexadecimal character strings that contain an IPv4 address. Add the strings to the TERMINAL class.

       For example, the IP address 163.97.227.17 is translated to A361E311. To allow all addresses in the 163.97.227.17 subnet, code the following statement:

       ```
       RDEFINE TERMINAL A361E3* UACC(READ)
       ```

    2. Ensure that login user IDs have READ access to the TERMINAL profile that includes their client system IP address.
  - If your RACF SETROPTS options are TERMINAL(READ), then all terminals are allowed access to your system and you do not need to add extra resource definitions to your RACF database.

- To establish SERVAUTH access, instead of terminal access, for IPv4 connection partners, specify PORTOFENTRY4 SERVAUTH in the FTP.DATA file. The FTP server will use the UNIX System Services _poe() service to identify the control socket as the port of entry.
- To establish SERVAUTH access for IPv6 connection partners, you do not need to do anything; IPv6 connection partners automatically establish SERVAUTH access. If the IPv6 connection partner is not in a network access security zone, the _poe() service does not pass a port of entry resource name and the port of entry is not checked. For information about network access security zones, see "Network access control" on page 140.

  For IPv4 and IPv6 users with either terminal or SERVAUTH access, you can optionally restrict access to DATASET resources during the login session by adding WHEN(TERMINAL=...) or WHEN(SERVAUTH=...) conditions to DATASET resource profiles in RACF.

Results

When you are finished, access to the FTP server is controlled based on the client's port of entry.

(Optional) Steps for controlling user access to the z/OS UNIX file system

FTP uses the resource profile EZB.FTP.*sysname.ftpdaemonname*.ACCESS.HFS in the SAF SERVAUTH class to control access to the z/OS UNIX file system. If you do not control access to this profile, then all users can access your z/OS UNIX file system.

Before you begin

If the FTP.DATA file for the server specifies STARTDIRECTORY HFS and the user is not permitted to the SERVAUTH class profile, FTP makes the TSO user ID the starting directory.

You must have the authority to issue the necessary RACF commands.

Procedure

Perform the following steps to control access to the z/OS UNIX file system:

1. Define the profile for the FTP user access to the z/OS UNIX file system.

   The profile has the following form:

   ```
   RDEFINE  SERVAUTH   EZB.FTP.sysname.ftpdaemonname.ACCESS.HFS
   ```

   For example, the profile name for FTP daemon FTPD running on system MVSA is the following name:

   ```
   EZB.FTP.MVSA.FTPD1.ACCESS.HFS
   ```

   Tip: The profile name can contain wildcard values as allowed by the security product. All security-product rules (for example wildcards, PROTECTALL, and so on) apply. For example, if all systems will use the same access list and RACF generic profile checking is active for the SERVAUTH class, you could use the following profile name:

   ```
   EZB.FTP.*.FTPD.ACCESS.HFS
   ```

2. Permit the user IDs that require access to the z/OS UNIX file system to the profile:

   ```
   PERMIT EZB.FTP.sysname.ftpdaemonname.ACCESS.HFS CL(SERVAUTH)
          ID(ftpuser)
   ```

3. Issue the following command to activate the RACF SERVAUTH class, if it is not already activated:

   ```
   SETROPTS CLASSACT (SERVAUTH)
   ```

4. Take one of the following actions:

   - RACLIST the SERVAUTH class, if this is a new profile:

     ```
     SETROPTS RACLIST (SERVAUTH)
     ```

- Refresh the SERVAUTH class, if you have changed an existing profile:

```
SETROPTS RACLIST (SERVAUTH) REFRESH
```

Results

When you are finished, only certain users will be able to access the z/OS UNIX file system.

Preventing exploitation of your FTP server

Your FTP server can be used by a client for disruptive purposes. A client can use your server to send random data to other servers, or a client can request that your server be the passive server in a three-way transfer.

Any FTP client that is in PROXY mode with your FTP server can establish a data connection to any server that is listening to a port. This situation could be very disruptive to that server, because the client could then send a very large amount of unexpected data to it. Any malicious FTP client can attack or disrupt the server in a normal server-to-client connection by making the FTP server send a large amount of data to another application server that is listening to a specific port. Because the client itself is not sending the disruptive data, it is difficult to identify the client that is causing the problem. Use the PORTCOMMAND, PORTCOMMANDPORT and PORTCOMMANDIPADDR statements in FTP.DATA to prevent your server from being used in this way.

Table 31. PORTCOMMAND scenarios

When you want your server to...	Code the following statements in the server's FTP.DATA	Comments
Reject all PORT or EPRT commands	PORTCOMMAND REJECT	If you disable the PORT or EPRT commands, then you prevent your server from being used to send random data to other servers. However, your server loses some ability to transfer data in PROXY mode. If a client sends a PORT or EPRT command to your server to set up a proxy transfer, your server will reject the command and the proxy transfer will fail. If your client is not firewall friendly, and it does not implement the default port number and IP address for data transfer, that client cannot transfer files to and from your server.
Reject all PORT or EPRT commands that specify well-known ports (port numbers less than 1024)	PORTCOMMANDPORT NOLOWPORTS	When you specify this combination, your server cannot be used to send random data to servers listening on well-known ports. However, a rogue client can use your server to send random data to servers listening on other ports. The server still supports data transfer in PROXY mode.
Reject all PORT or EPRT commands that specify an IP address other than the client's own IP address.	PORTCOMMANDIPADDR NOREDIRECT	When you specify this combination, a client can request data transfer in PROXY mode only between your server and a server on its own IP address. Transfers between client and server are not affected.

Table 31. PORTCOMMAND scenarios (continued)		
When you want your server to...	**Code the following statements in the server's FTP.DATA**	**Comments**
Reject all PORT or EPRT commands that specify an IP address other than the client's own IP address or port numbers that are well known.	PORTCOMMANDPORT NOLOWPORTS PORTCOMMANDIPADDR NOREDIRECT	When you specify this combination, a client can request data transfer in PROXY mode only between your server and a server that is on its own IP address; the port numbers cannot be well known. The client cannot use PROXY mode to send random data to a server that is on its own IP address and listening to a well-known port.

Your FTP server can also be used as a passive server in a three-way transfer. When a client sends a PASV or EPSV command to the server, the server opens a listening data socket. This socket is similar to the listening socket associated with the well-known port, in the sense that any application that knows its IP address and port number can connect to it (not just the client that sent the PASV or EPSV command). The client can exploit this situation to initiate a three-way data transfer, which is a data transfer between two servers. The client sends PASV to one server followed by PORT to the other. The client sets the PORT command IP address and port number to the information it gets from the PASV reply, and the second server connects to the IP address and port number specified in the PORT command, connecting the two servers. The next data transfer command causes data to move directly between the two servers. The client can also use the EPSV and EPRT commands to set up the three-way data transfer.

Three-way transfers are supported functions in the FTP protocol, but you might not want to allow your server to participate in three-way transfers. To prevent your server from being the passive server (the server that receives the PASV or EPSV command) in a three-way data transfer, code PASSIVEDATACONN NOREDIRECT in the server's FTP.DATA file. This directs the server to verify that the data connection comes from the IP address where the original FTP client is (the client that sent the PASV or EPSV command). If that is not where the data connection originates, the server closes the data socket and the next data transfer command fails.

To completely disallow the use of your FTP server in three-way transfers, code the PASSIVEDATACONN statement as described in the preceding paragraph, and the PORTCOMMANDIPADDR NOREDIRECT statement or PORTCOMMAND REJECT statement described in Table 31 on page 671.

(Optional) Assigning password phrases to user IDs that are used to log in to the FTP server

You can enable users to log in to the FTP server with password phrases by assigning password phrases to the user IDs that are used to log in to the FTP server.

Guideline: Use the RACF ALTUSER command to assign a password phrase to a user ID. See assigning password phrases in the z/OS Security Server RACF Security Administrator's Guide for more information.

Rules:

- Do not assign a character or combination of characters to a password phrase that your FTP clients cannot support.
- The maximum length of a password phrase is 100 characters.

Tips:

- Use alphanumeric characters to create password phrases to minimize translation problems when the client and server are using different ASCII code pages.
- If you plan to use the z/OS FTP client in batch mode, the password phrase and optional user data must fit on a single line of the batch file.

Results:

- When you assign a password phrase to a user ID, a user can log in with that user ID by using either the password or the password phrase.
- A user can change the password phrase when they log in to FTP.

Restrictions:

- The password phrase that a user enters to log in to the z/OS FTP server has restrictions beyond those that are enforced by your security product and those that are enforced by the optional ICHPWX11 user exit. The password phrase must not contain the following characters that have special meaning to the z/OS FTP server:
  - NULL (X'00')
  - Slash (/)
  - Colon (:)
  - Carriage return (<cr>)
  - Line feed (<lf>)
  - Interpret as command (<IAC> or X'FF')
  - Telnet command characters (X'FB' - X'FE')

  The z/OS FTP server translates all passwords that it receives during a session into EBCDIC before it passes them to the security product and to the FTCHKPWD exit routine.
- A user will not be able to log in to the z/OS FTP server using a password phrase if you assign a character to the password phrase that the server cannot translate from ASCII or UTF-8 to EBCDIC. The user will also not be able to log in if you assign a character that the server translates into an EBCDIC character that is different than the character that you assigned. Untranslatable characters and inconsistent translations can occur if the client and server are using different code pages, or if the character is outside the normal range of printable characters.
- The z/OS FTP server supports quotation marks in password phrases, but your FTP client might not support the use of quotation marks in password phrases.
- The password phrase must not contain leading blanks or trailing blanks.
- The maximum length of a password phrase is 100 characters.
- When you configure the z/OS FTP server for anonymous FTP, the following restrictions apply:
  - You cannot specify a password phrase instead of a password as an FTP daemon start option.
  - You cannot code a password phrase instead of a password on the ANONYMOUS statement in FTP.DATA.

  If the server is configured to prompt anonymous users for a password, the user can log in with either the password or the password phrase that is assigned to the anonymous user ID.

Defining environment variables for the FTP server (optional)

The FTP server optionally uses environment variables to identify the translate table data sets to be used for the control and data connections. The FTP server uses the following environment variables to override a default naming convention.

Using _FTPXLATE_name for translation

In your FTP.DATA file, you can use the CCXLATE or XLATE statements to specify a name that corresponds to a particular data set that is to be used for the initial translate tables for the control or data connections.

Rules:

- CCXLATE is ignored if EXTENSIONS UTF-8 or CTRLCONN is coded in FTP.DATA.
- XLATE is ignored if SBDATACONN is coded in FTP.DATA.

FTP searches for an environment variable defined as _FTPXLATE_*name=fully_qualified_dsn* with the following specifications:

name
The XLATE or CCXLATE parameter value you coded in FTP.DATA. You can code 1 to 8 alphabetic or numeric characters.

fully_qualified_dsn
A fully qualified MVS data set name or z/OS UNIX file name.

Requirements:

- You must use uppercase letters for the *name* value of the _FTPXLATE_*name* environment variable.
- The *name* value must follow the rules for an MVS data set qualifier if you do not define the _FTPXLATE_*name* environment variable.

FTP identifies the data set corresponding to the *name* value:

- If the environment variable is defined, FTP uses the data set name that is defined by the environment variable.
- If the environment variable is not defined, FTP uses the data set name *hlq*.*name*.TCPXLBIN.

You can also issue the SITE XLATE= command from any client to set the translate tables for the data connection for that particular FTP session. The FTP server searches for an environment variable called _FTPXLATE_*name* and uses the data set name that _FTPXLATE_*name* defines. If the _FTPXLATE_*name* environment variable does not exist, the server searches for a data set called *hlq*.*name*.TCPXLBIN.

Using TZ and other UNIX environment variables

Use the ENVAR runtime option in your FTPD start procedure to set environment variables for the FTP server. For information about using the ENVAR runtime option to set environment variables, see z/OS XL C/C++ Programming Guide. The following example shows how to use ENVAR to specify environment variables in your FTPD started procedure:

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//       PARM=('POSIX(ON) ALL31(ON)',
//            'ENVAR("TZ=EST")/&PARMS')
```

Using _BPX_JOBNAME for similar job names

If you want all FTP forked tasks to have similar job names, set the _BPX_JOBNAME environment variable. For example, for WorkLoad Manager (WLM), accounting, and isolation of syslogd messages, you might not want each FTP user that is logged in to have its user ID for a job name.

The following example shows how to use the _BPX_JOBNAME environment variable to assign the job name FTPD to all FTP forked tasks:

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//       PARM=('POSIX(ON) ALL31(ON)',
//            'ENVAR("_BPX_JOBNAME=FTPD"',
//            '"TZ=EST")/&PARMS')
```

Requirement: If you activate the SAF class APPL and define a resource profile in that class for the job name resource that you specify with the _BPX_JOBNAME environment variable, you must grant user IDs that log in to FTP at least READ access to that resource.

Using _BPXK_SETIBMOPT_TRANSPORT for an affinity to a specific stack

If you want to ensure that FTP has an affinity to a particular TCP/IP stack, use the _BPXK_SETIBMOPT_TRANSPORT environment variable. For more information, see “Generic server versus server with affinity for a specific transport provider” on page 46.

The following example shows how to use the _BPXK_SETIBMOPT_TRANSPORT environment variable to set the FTP server to have an affinity to TCPIPOE:

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//      PARM=('POSIX(ON) ALL31(ON)',
//           'ENVAR("_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE"',
//           '"TZ=EST")/&PARMS')
```

Configuring FTP with multiple TCP/IP stacks

Prior to configuring the FTP server with multiple TCP/IP stacks, review "Considerations for multiple instances of TCP/IP" on page 45.

The FTP server can be configured as a server with affinity to a specific transport provider or as a generic server.

To configure the FTP server to have affinity to a specific transport provider, take the following actions:

- Code the _BPXK_SETIBMOPT_TRANSPORT keyword in the FTP cataloged procedure. The example below sets the FTP server to have an affinity to TCPIPOE.

```
//FTPD   PROC MODULE='FTPD',PARMS=''
//FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
//      PARM=('POSIX(ON) ALL31(ON)',
//           'ENVAR("_BPXK_SETIBMOPT_TRANSPORT=TCPIPOE"',
//           '"TZ=EST")/&PARMS')
```

- Reserve ports 21 and 20 for the FTP server in PROFILE.TCPIP as follows:

```
         PORT
             21 TCP FTPD1          ; FTP server control port
             20 TCP OMVS NOAUTOLOG ; FTP server data port
```

To configure the FTP server as a generic server, reserve ports 21 and 20 for the FTP server in PROFILE.TCPIP on all transport providers. The FTP server will detect when new transport providers are activated and attempt to bind to port 21. If this port is not reserved for the FTP server, the FTP server will end with the following message:

```
EZYFT13E bind error : EDC5111I Permission denied
```

Configuring TCPIP.DATA for FTP

The following five statements are used by the FTP server:

DATASETPREFIX
Specifies HLQ for dynamic allocation

DOMAINORIGIN
Specifies the domain name to be appended to host name

HOSTNAME
Specifies the TCP host name

LOADDBCSTABLES
Specifies the DBCS tables used by the client and server

MESSAGECASE
Specifies the case that messages should be displayed in

See Chapter 2, "IP configuration overview," on page 11 for information about TCPIP.DATA or see z/OS Communications Server: IP Configuration Reference for information about these statements.

Configuring FTP.DATA

The FTP.DATA data set is optional. The FTP daemon looks for this data set during initialization, using the first file it finds in the following search order:

1. A data set specified by the //SYSFTPD DD statement
2. *ftpserve_job_name*.FTP.DATA
3. /etc/ftp.data
4. SYS1.TCPPARMS(FTPDATA)
5. *hlq*.FTP.DATA data set

It is not necessary to include all statements in the FTP.DATA data set. Include the statements only if the default value is not what you want, because the default will be used for any statement not included in the FTP.DATA data set.

To pick up changes made in the FTP.DATA data set, the FTP server must be stopped and restarted. Some FTP server parameters can be changed during an FTP session by the client issuing the SITE subcommand. See z/OS Communications Server: IP User's Guide and Commands for more information. The FTP client has an FTP.DATA data set which can also be used to change the defaults for the FTP client local site parameters. See the z/OS Communications Server: IP User's Guide and Commands for more information about using the FTP.DATA data set for the FTP client local site parameters.

Optionally configuring user-level server options using FTPS.RC

The default values for the site parameters are coded in the server FTP.DATA. These SITE defaults apply to all login sessions to the server. You can customize settings for a specific user or group of users by creating an FTPS.RC configuration data set containing FTP commands specific to that login session. This file may contain a series of CWD and SITE commands. See z/OS Communications Server: IP User's Guide and Commands for information about these commands.

The FTP server uses the following search order to find the MVS data set or z/OS UNIX file:

1. tso_prefix.FTPS.RC
2. userid.FTPS.RC
3. $HOME/ftps.rc

Data set attributes

Data set attributes play a significant role in FTP performance. If your environment permits, tune both BLOCKSIZE and LRECL according to the following recommendations:

- Use half a track as the block size.
- For IBM 3380 DASD, use 23424 as the block size with an LRECL of 64 bytes.
- For IBM 3390 DASD or IBM9334, use 27968 as the block size with an LRECL of 64 bytes.
- Use FB as the data set allocation format.
- Use cached DASD controllers.
- If your environment permits, use a preallocated data set for FTP transfers into MVS.

The following configuration data statements apply to FTP server's allocation of data sets.

- AUTOMOUNT
- AUTORECALL
- BLKSIZE
- BUFNO
- CONDDISP
- DATACLASS

- DCBDSN
- DIRECTORY
- DSNTYPE
- EATTR
- LRECL
- MGMTCLASS
- MIGRATEVOL
- PDSTYPE
- PRIMARY
- RECFM
- RETPD
- SECONDARY
- SPACETYPE
- STORCLASS
- UCOUNT
- UMASK
- UNITNAME
- VCOUNT
- VOLUME

See z/OS Communications Server: IP Configuration Reference for more detailed information about these keywords.

Some of these allocation variables might provide duplicate information. FTP passes all variables that are specified to the z/OS dynamic allocation function to determine which of the specifications take precedence, except for the following exceptions:

- If the data set organization is physical sequential, directory blocks are not sent.
- If neither primary nor secondary space quantities are specified, the allocation units value is not sent.

For example, the model DCB (DCBDSN) might have a record format (RECFM) that differs from the record format specified by a data class and from the one explicitly specified by the client. The order of precedence for dynamic allocation variables are as follows:

1. Any FTP.DATA statements or SITE parameters explicitly specified or in effect by default.
2. Any attributes picked up from the model DCB and not otherwise explicitly specified.
3. Any attributes picked up from the data class and not previously derived from 1 and 2 above.
4. Any system allocation defaults.

Specifying attributes for new MVS data sets

When allocating new data sets, there are two methods you can use to specify the data set attributes. You can customize the data set attributes for your login session using the SITE command, or you can configure the data set attributes for logging in to your server using statements in FTP.DATA. Or, if your system programmer has used the Storage Management System to group together default attributes into named classes, you can specify those class names on the DATACLASS, STORCLASS, and MGMTCLASS statements.

Dynamic allocation

The FTP server allows a client program to dynamically allocate a new physical sequential data set, a partitioned data set (PDS), or a partitioned data set extended (PDSE), for the purpose of transferring data

to be written to that data set. The following optional allocation variables can be used to override and turn off the defaults that affect the allocation of the data set.

Variable
FTP.DATA statement

allocation units
SPACETYPE

blocksize
BLKSIZE

data class
DATACLASS

directory blocks
DIRECTORY

data set name type (physical sequential data sets only)
DSNTYPE

extended attributes
EATTR

logical record length
LRECL

management class
MGMTCLASS

model DCB values
DCBDSN

PDS type
PDSTYPE

primary space
PRIMARY

retention period
RETPD

secondary space
SECONDARY

space type
SPACETYPE

unit count
UCOUNT

volume count
VCOUNT

record format
RECFM

retention period
RETPD

storage class
STORCLASS

unit
UNITNAME

volume serial number or list
VOLUME

Some of these allocation variables might provide duplicate information. For example, the model DCB might have a record format (RECFM) that differs from the record format specified by a data class and from the one explicitly specified by the client. FTP passes all variables that are specified to dynamic

allocation and lets it determine which of the specifications take precedence. The following list describes the exceptions to that policy:

- If neither the primary nor secondary space quantity is specified, the allocation units value is not sent.
- If the data set organization is physical sequential, directory blocks specification is not sent.
- Otherwise, all variables are sent to dynamic allocation where the order of precedence is:
  1. Any FTP.DATA statements or SITE parameters explicitly specified or specified by default
  2. Any attributes picked up from the model DCB and not otherwise explicitly specified
  3. Any attributes picked up from the data class and not previously derived from 1 or 2
  4. Any allocation defaults

Storage Management Subsystem

You can specify one or more of the following Storage Management Subsystem (SMS) classes to manage characteristics that are associated with or assigned to data sets.

- Data class is an SMS construct that an installation can define to control data set allocation attributes used by SMS for the creation of data sets. An installation can override all or part of an SMS DATA CLASS definition by using FTP.DATA statements. Note that there is an order of precedence for dynamic allocation. (See "Data set attributes" on page 676 for more information on the precedence.) The fields listed are available attributes that serve as a template for allocation. Each is optional and is overridden by any explicit specification of FTP-allocation variables. Except for the DSNTYPE and EATTR fields, the listed variables can be overridden by a model DCB (DCBDSN).

  Variable
  FTP.DATA statement

  directory blocks
  DIRECTORY

  data set name type
  DSNTYPE

  extended attributes
  EATTR

  logical record length
  LRECL

  primary space
  PRIMARY

  record format
  RECFM

  retention period
  RETPD

  secondary space
  SECONDARY

  pds type
  PDSTYPE

  Note: If either primary or secondary space is explicitly specified, the primary and secondary values from data class are not used.
- Management class (MGMTCLASS) is an SMS construct that determines DFHSM action for data set retention, migration, backup, and release of allocated but unused space. Management class replaces and expands attributes that otherwise would be specified. That is, management class might override any other specification of retention period.
- Storage class (STORCLASS) is a list of storage performance and availability services requests for an SMS-managed data set that SMS attempts to honor when selecting a volume or volumes for the data

set. It might conflict with an explicit specification of volume and unit. If storage class is used, volume and unit should not be specified.

Translation of data

Selecting an appropriate translate table for conversion of data from host to network format, and from network to host format, will ensure that data read from or written to the z/OS system are in correct format. The following statements apply to translation of data for the FTP server. See z/OS Communications Server: IP Configuration Reference for more information on these statements. The statements are:

- ASATRANS
- CTRLCONN
- DBSUB
- ENCODING
- EXTENSIONS UTF8
- MBDATACONN
- MBSENDEOL
- MBREQUIRELASTEOL
- SBDATACONN
- SBSENDEOL
- SBSUB
- SBSUBCHAR
- UCSHOSTCS
- UCSSUB
- UCSTRUNC
- UNICODEFILESYSTEMBOM

z/OS UNIX named pipes

FTP can transfer data to and from z/OS UNIX System Services named pipes. The following statements apply to the creation of named pipes:

- UMASK
- UNIXFILETYPE

The following statements apply to the transfer of data to and from named pipes:

- CONDDISP
- FIFOIOTIME
- FIFOOPENTIME

For more information about these FTP.DATA data set statements, see z/OS Communications Server: IP Configuration Reference. For information about using z/OS UNIX System Services named pipes, see z/OS Communications Server: IP User's Guide and Commands.

FTP code page conversion

Code page conversion must be performed for:

- FTP commands and replies sent over the control connection
- Data transferred over the data connection

FTP uses the iconv function to establish ASCII-to-EBCDIC and EBCDIC-to-ASCII translate tables for the control connection. The default network transfer code page for the control connection is 7-bit ASCII.

In addition, FTP maintains support for the use of translate tables by the CONVXLAT utility. After a user has logged in, a SITE subcommand can be used to change the code page being used on the control connection.

FTP uses the iconv function to establish network transfer to file system and file system to network transfer translate tables for the data connection. In addition, FTP maintains support for the use of translate tables by the CONVXLAT utility.

Note: Using iconv conversion to retrieve EBCDIC data that was created with CONVXLAT-generated conversion tables could result in data corruption due to possible conversion table differences.

After a user has logged in, SITE and LOCSITE subcommands can be issued to change the translation tables being used for single-byte translation.

Code page conversions for the control connection

For the control connection, FTP generally uses ASCII for the network code page, as specified in the FTP RFCs. For the host/ASCII conversion for the control connection, FTP uses either iconv() or the support for single-byte translation tables. However, when EXTENSIONS UTF8 is coded in FTP.DATA, FTP starts the connection in 7-bit ASCII and negotiates a switch to UTF-8 encoding of the control connection, as described in RFC 2640. FTP uses iconv() for the host/UTF-8 conversion.

Priority

The priority for establishing the conversion tables used for the control connection is:

1. FTP start parameter (FTP client only)
2. EXTENSIONS UTF8 coded in FTP.DATA
3. CTRLCONN or CCXLATE keyword in FTP.DATA
4. Search order used to locate a TCPXLBIN data set:
    a. Original jobname.SRVRFTP.TCPXLBIN
    b. *hlq*.SRVRFTP.TCPXLBIN
    c. Original jobname.STANDARD.TCPXLBIN
    d. *hlq*.STANDARD.TCPXLBIN
5. 7–bit ASCII
6. Internal (hard-coded) 7–bit tables

Code page conversions for the data connection

For the transfer of data on the data connection, FTP supports:

- All single-byte conversions available through iconv. For example, iconv supports conversions between IBM-1047 and IBM-850, so conversions between IBM-1047 and IBM-850 are available for data transfer.
- UNICODE conversions available through iconv: UTF-8, UTF-16, UTF-16BE, and UTF-16LE for network transfer, and UTF-8 and UTF-16 for file storage.
- Multibyte conversions for the Chinese standard GB18030 using code page IBM-5488 with code page IBM-1388 or UTF-8, as well as certain double-byte character set (DBCS) code page pairs that are equivalent to supported DBCS languages.
- Both single-byte and double-byte data conversions are supported with the translate tables provided with TCP/IP or generated by the CONVXLAT utility.

Priority for single-byte conversions

How to build single-byte translate tables for converting network transfer data and file system data is determined by the following priority:

- SYSFTSX DD statement in the startup procedure, where the named data connection is CONVXLAT-generated translate tables. The data set can be an MVS data set or a z/OS UNIX file.
- SBDATACONN or XLATE keyword in FTP.DATA.
- Search order to locate a TCPXLBIN data set, where the MVS data set contains CONVXLAT-generated translate tables:
  1. Original jobname.SRVRFTP.TCPXLBIN
  2. *hlq*.SRVRFTP.TCPXLBIN
  3. Original jobname.STANDARD.TCPXLBIN
  4. *hlq*.STANDARD.TCPXLBIN
- The same conversions established for the control connection.

Multibyte character sets (MBCS) support

MBCS support in the FTP server is provided for the following code page pairs:

Support for:	**TYPE command**	**File system code page**	**Network transfer code page**
Chinese standard GB18030	Not applicable	IBM-1388 or UTF-8	IBM-5488
BIG5	TYPE B 8	IBM-937	IBM-950 or BIG5
EUCKANJI	TYPE B 2	IBM-930	IBM-eucJP
JIS78KJ (JISROMAN	TYPE B 4 R	IBM-930	IBM-5053
JIS78KJ (ASCII	TYPE B 4 A	IBM-939	IBM-5055
JIS83KJ (JISROMAN	TYPE B 3 R	IBM-930	IBM-5052
JIS83KJ (ASCII	TYPE B 3 A	IBM-939	IBM-5054
KSC5601	TYPE B 6	IBM-933	IBM-949
SCHINESE	TYPE B 9	IBM-935	IBM-1381
SJISKANJI	TYPE B 1	IBM-930 or IBM-939	IBM-932 or IBM-eucJC
TCHINESE	TYPE B 7	IBM-937	IBM-948
UNICODE	Not applicable	UTF-8, UTF-16	UTF-8, UTF-16, UTF-16BE, and UTF-16LE

Rules:

- ENCODING must be specified as MBCS, either in FTP.DATA or on a SITE command.
- The data type must be ASCII.
- The file structure must be FILE (not RECORD), and the transfer mode must be STREAM (not BLOCK or COMPRESS).
- The FTP file type must be SEQ (not JES or SQL).
- If the file is transferred to or from an MVS data set, the record format of the data set must be V, VB, or U.
- If a file is transferred outbound and is an MVS data set with record format V or VB, the request for RDWs is not allowed.
- Translation of ASA or machine control characters is not allowed.
- MBCS can be used as a migration path for the DBCS languages listed that have an associated TYPE B *x* command.

Restriction: The DBCS languages can be migrated to the MBCS support only if they do not use the following parameters on the TYPE B *x* command:

S A
SOSI ASCII characters X'1E' and X'1F' in the ASCII data stream

S E
SOSI EBCDIC characters X'0E' and X'0F' in the ASCII data stream

S S
SOSI SPACE characters X'20' and X'20' in the ASCII data stream

N
No SOSI characters in the ASCII data stream and none written to the file system

Master catalog access

FTP uses the IGGCSI00 function to request catalog processing. This accesses both the user and master catalog. Users require READ access to the master catalog as well as their own user catalog.

Customizing FTP message catalogs

FTP messages and replies are contained in two z/OS UNIX message catalogs as follows:

- ftpdmsg.cat

  Contains messages that the FTP daemon, server, and client issue.
- ftpdrply.cat

  Contains replies that the server sends to the client.

If messages in either of these catalogs need to be modified, the timestamp that is contained in the shipped level of the FTP catalog must be preserved in the modified catalog.

This timestamp is included within and is unique to each catalog. When FTP (client or daemon) is started, FTP verifies that this timestamp matches the timestamp that it expects. This prevents FTP from presenting the wrong message when the z/OS UNIX message catalogs and FTP are not synchronized.

When you apply a service update to FTP load modules that requires a service update to a catalog, you must install both at the same time. Otherwise, FTP will report an error and use default messages instead of messages from the catalog.

If you are using a modified FTP catalog, the catalog that matches the service level of FTP load modules needs to be updated with your local modifications, and the timestamp of the catalog must be preserved.

Steps for creating a message catalog from the shipped catalog and preserving its timestamp

After using these steps you can then update the file with your local modifications and create a new FTP catalog with the preserved timestamp. Perform all of these steps from the z/OS UNIX shell; the commands indicated are z/OS UNIX commands.

Before you begin

This example assumes that the FTP code and catalog are at the correct levels and that only local customization to the catalog is to be performed. Also, if you customize a catalog, IBM Service personnel might require that you use the shipped level of the catalog to recreate and diagnose a reported problem.

Procedure

Perform the following steps to create a file from the shipped catalog and preserve its timestamp:

1. Copy the official z/OS UNIX catalog that is shipped with the release or service to a backup file.

```
cp  /usr/lpp/tcpip/lib/nls/msg/C/ftpdmsg.cat  /tmp/ftpdmsg.cat.backup
```

2. Using the **dspcat** command, convert a copy of the backup catalog that you just copied to a copy that you can edit.

   This is the file that you need to update to preserve the timestamp, and that you will update to support any local user message changes.

```
dspcat -t -g  /tmp/ftpdmsg.cat.backup  >/tmp/ftpdmsg.cat.copy
```

3. Change the first line in the catalog from a comment to a z/OS UNIX **gencat** command, which enables the timestamp to be imbedded in the catalog when it is rebuilt.

   a) Edit the file to be updated.

```
oedit /tmp/ftpdmsg.cat.copy
```

   b) Change the first line comment to add a **gencat** command to preserve the timestamp when the directory is built.

   The first line in the file will be similar to the following line:

```
The time stamp of catalog /tmp/ftpdmsg.cat.backup is: 2006 095 20:30 UTC
```

   Replace the leading text on the line with the **gencat** subcommand **$timestamp** as follows:

```
$timestamp 2006 095 20:30 UTC
```

   If this step is omitted and the original line is left in the catalog, when an attempt is made to generate a catalog from this file, you will see a message similar to the following message:

```
FSUM5108 gencat: Invalid message number.
```

   c) Save the file.

4. Update the catalog (/tmp/ftpdmsg.cat.copy) with any local modifications and save the file.
5. Build a new and customized catalog using the z/OS UNIX **gencat** command, and save a copy of the shipped level of the catalog.

```
gencat   /tmp/ftpdmsg.cat   /tmp/ftpdmsg.cat.copy
```

   The correct response is:

```
FSUM5105 gencat: Message catalog generated normally.
```

6. Browse the new catalog and verify that the timestamp from step “3” on page 684 matches what is in the file.

```
obrowse /tmp/ftpdmsg.cat
```

   The first record contains the time stamp (*yyyy ddd hh:mm* UTC). For example:

```
…2006 095 20:30 UTC
```

7. Replace the official z/OS UNIX catalog that is shipped with the release with the updated catalog that you created in step “5” on page 684.

Results

You know you are done when you test the catalogs to verify correct synchronization by performing the following actions:

1. Start the FTP server and inspect the syslog output (console if no syslog is running). Message EZYFS30W should not appear.
2. Using the FTP client, connect to the FTP server and inspect the SYSPRINT output. Message EZYFS31W should not appear.

For more information about the **dspcat** and **gencat** utilities, see z/OS UNIX System Services Command Reference.

Accounting

The following parameters apply to SMF data:

- SMF
- SMFAPPE
- SMFDEL
- SMFEXIT
- SMFJES
- SMFLOGN
- SMFREN
- SMFRETR
- SMFSQL
- SMFSTOR

See z/OS Communications Server: IP Configuration Reference for more information on these statements.

Configure the FTP server for SMF (optional)

The FTP server can write SMF type 118 (X'76') or type 119 (X'77') records to record transactions made by the FTP server. SMF records can be written for the following commands:

- APPE (append)
- DELE (delete)
- RNTO (rename)
- RETR (retrieve)
- STOR (store)
- STOU (store unique)

Information about the previous commands can be recorded for:

- FTP server running in normal data transfer mode (FILETYPE=SEQ)
- FTP server running remote job submission (FILETYPE=JES)
- FTP server running Structured Query Language (SQL) queries (FILETYPE=SQL)
- Any combination of SEQ, JES, and SQL

For commands involving data transfer (APPEND, RETR, STOU or STOR) an SMF record will be written for both successfully and unsuccessfully completed data transfer commands which have begun data transfer. For data transfer commands which have completed unsuccessfully, the byte count of transmission field will contain the number of bytes transferred before the failure, and the recent server reply field will contain the 3-digit error reply code sent to the client. See the information about type 118 records and type 119 records in z/OS Communications Server: IP Programmer's Guide and Reference to find the particular offsets for the record type being used.

The FTP server can also write SMF records when a login attempt fails.

The capability also exists for a user-written exit routine to get control before the SMF records are written. See “Configuring the optional FTP user exits” on page 708 for more information.

If you want the FTP server to write SMF type 118 (X'76') or type 119 (X'77') SMF records, you must include at least one of the SMF subtype statements (SMF, SMFAPPE, SMFDEL, SMFLOGN, SMFREN, SMFRETR, or SMFSTOR) in the FTP.DATA data set.

If SMF subtype statements are not coded in the FTP.DATA data set, no SMF records are written by the FTP server.

For more information about enabling SMF TYPE 119 records for FTP Server, see SMF (FTP server) statement in z/OS Communications Server: IP Configuration Reference.

Customizing Transport Layer Security and Kerberos security

The following terms apply to Transport Layer Security (TLS) and Kerberos.

Integrity protected, data integrity, or data authentication
Indicates an algorithm is applied to the data being transferred, which modifies the data such that the receiving program can verify the data was not modified or changed during the transfer.

Privacy protected
Indicates an algorithm is applied to the data being transferred, which encrypts or scrambles the data such that only the receiving program can use a special key to decrypt or unscramble the data to its original format. The original data cannot be seen or interpreted while the data is in transit.

Raw
Indicates data is transmitted without being modified by any encryption or data integrity algorithms.

Encipher or cipher algorithm
Data being transferred is encrypted, integrity protected, or both. This term does not imply which algorithm is used and does not imply the data is encrypted.

Steps for customizing the FTP server for TLS

You can customize the FTP server for TLS using AT-TLS.

Before you begin

Understand the following information:

- The FTP server must be enabled to support both TLS and Kerberos. Some of the configuration statement settings apply to both TLS and Kerberos and affect the behavior of both.
- To support TLS, the FTP server must be configured to use Application Transparent TLS (AT-TLS) as a controlling application. For more information about AT-TLS, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

  Note: The use of TLS requires the use of X.509 digital certificates. For more about certificates and key ring databases, see Appendix B, “TLS/SSL security,” on page 1329.

Requirement: AT-TLS requires Policy Agent to be configured, and the TCP/IP stack to be enabled for AT-TLS. To configure AT-TLS, see “Configuring the server system” on page 1126.

Procedure

Perform the following steps to customize the FTP server for TLS:

1. Decide what level of RFC 4217, *On Securing FTP with TLS*, that you want the server to support.
   - To have the server support *On Securing FTP with TLS* at the Internet draft level, code the following statement in the server's FTP.DATA configuration file:

     ```
     TLSRFCLEVEL DRAFT
     ```

     This is the default. The z/OS FTP server has supported TLS security at this level since V1R2. Code this statement in FTP.DATA to maintain this level of support.
   - To have the server support *On Securing FTP with TLS* at the RFC 4217 level, code the following statement in the server's FTP.DATA configuration file:

     ```
     TLSRFCLEVEL RFC4217
     ```

The RFC *On Securing FTP with TLS* was published as RFC 4217 in October, 2005. The RFC differs from the Internet draft in its description of the AUTH, CCC, and REIN commands. RFC 4217 is less restrictive than the Internet draft regarding when the AUTH and CCC commands can be sent to the server, and more explicit about the details of the server REIN implementation. For more information, see RFC 4217.

2. Code the following statement in the server's FTP.DATA configuration file to enable the server for TLS:

```
EXTENSIONS AUTH_TLS
```

3. Decide what level of authentication you will use for TLS sessions:
   - Server authentication only
   - Client authentication level 1
   - Client authentication level 2
   - Client authentication level 3

   For more information about server authentication and client authentication, see "Secure Socket Layer overview" on page 1329.
4. Create the server key ring database and add the certificates you will need to the server key ring database.

   Every TLS session handshake includes server authentication, so you must always add a certificate for this server to the server key ring database. If a server certificate is self signed, you must also export that certificate to the key ring databases of those clients that will log in using TLS. If a server certificate is signed by a certificate authority (CA), the CA certificate used to sign the server certificate needs to be in the client key ring databases, rather than the server certificate. For more information about server authentication, see "Server authentication " on page 1329.

   If you are using client authentication and self-signed certificates, you must import the client certificates into the server key ring database. If a client certificate is signed by a CA, the CA certificate used to sign the client certificate needs to be in the server key ring database, rather than the client certificate. For more information, see "Client authentication" on page 1330.
5. Code TLSMECHANISM ATTLS in the FTP.DATA. This is the default setting.
6. Decide whether clients logging in to this server should be required to use the TLS protocol.

   The default is to allow the client to decide whether to use TLS. This setting is customized by using the SECURE_FTP configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   To allow the client to decide whether to use TLS, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_FTP ALLOWED
```

   This is the default setting, and indicates:
   - If the server is enabled for TLS only, clients must either log in using TLS, or with no security mechanism.
   - If the server is enabled for Kerberos only, clients must either log in using Kerberos, or with no security mechanism.
   - If the server is enabled for both TLS and Kerberos, clients can log in using TLS, Kerberos, or with no security mechanism.

   To require that clients log in using a security mechanism, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_FTP REQUIRED
```

   This setting indicates:
   - If the server is enabled for TLS only, clients must log in using TLS.

- If the server is enabled for Kerberos only, clients must log in using Kerberos.
- If the server is enabled for both TLS and Kerberos, clients must log in using either TLS or Kerberos.

7. If you do not want to use client authentication, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_LOGIN NO_CLIENT_AUTH
   ```

   This is the default.

   If you do want to use client authentication, the following levels of client authentication are possible:

   - Level 1 authentication is performed by system SSL. The client passes an X.509 certificate to the server. To pass authentication, the Certificate Authority that signed the client certificate must be considered trusted by the server. To use level 1 client authentication, code the following statement in the server's FTP.DATA configuration file:

     ```
     SECURE_LOGIN REQUIRED
     ```

   - Level 2 authentication provides level 1 authentication, and additionally requires that the client certificate be registered with RACF (or another SAF compliant security product) and mapped to a user ID. The client certificate received during the SSL handshake is used to query the security product to verify that the certificate maps to a user ID known to the system prior to connection negotiation. To use level 2 client authentication, code the following statement in the server's FTP.DATA configuration file:

     ```
     SECURE_LOGIN VERIFY_USER
     ```

   - Level 3 authentication provides level 1 and 2 authentication. In addition, it provides the capability to restrict access to the server based on the user ID returned from RACF. If the SERVAUTH class of RACF is active and the server's port profile is defined, a connection is accepted only if the requester's user ID associated with the client certificate is defined in the server's port profile. To use level 3 client authentication, code the following statement in the server's FTP.DATA configuration file:

     ```
     SECURE_LOGIN VERIFY_USER
     ```

     Also, define the server's port profile in the SERVAUTH class of RACF.

   If you choose to use client authentication, you can also use the client certificate authentication process to eliminate the client login password prompt so that a client supplies only the login user ID to establish the session. The certificate received from the client must be registered in the security product and must be associated with the login user ID. You can use the RACDCERT ADD command to register and associate the certificate. If either the certificate is not registered or is not associated with the user ID, you will be prompted for a password.

   If you do not want to use the client authentication process to eliminate the client password prompt, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_PASSWORD REQUIRED
   ```

   This is the default.

   If you want to use the client authentication process to eliminate the client password prompt, along with your client authentication statement (either SECURE_LOGIN REQUIRED or SECURE_LOGIN VERIFY_USER), code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_PASSWORD OPTIONAL
   ```

8. Configure the AT-TLS policy for the FTP server.

   To configure AT-TLS, see “Configuring the server system” on page 1126.

   Requirements:

- The FTP server is a controlling application. For more information about controlling applications, see "Advanced application considerations" on page 1144.

  Code a TTLSEnvironmentAdvancedParms statement with the ApplicationControlled and SecondaryMap parameters; both parameters should specify the value On. The ApplicatonControlled parameter allows FTP to start and stop TLS security on a connection. The SecondaryMap parameter enables active or passive data connections to use the AT-TLS policy that is used for the control connection. You do not need to code any additional TTLSRule statements for the data connections.
- The FTP server requires that the HandshakeRole parameter with the value Server or ServerWithClientAuth be coded on the TTLSEnvironmentAction statement. If the SECURE_LOGIN statement is coded in FTP.DATA with the parameters REQUIRED or VERIFY_USER, the HandshakeRole parameter value must be ServerWithClientAuth.
- The TTLSRule statement for the FTP server requires the Direction parameter with the value Inbound.

A sample Policy Agent AT-TLS configuration showing the required policy configuration statements for AT-TLS is as follows:

```
   TTLSGroupAction  secure_ftp_server_group
   {
      TTLSEnabled On
   }
   TTLSEnvironmentAction secure_ftp_server_env
   {
      TTLSKeyringParms
      {
         Keyring FTPD/server-keyring-database
      }
      HandshakeRole Server       # When Secure_Login NO_CLIENT_AUTH is coded
      #HandshakeRole ServerWithClientAuth # When Secure_Login Required or Verify_User is coded
      TTLSEnvironmentAdvancedParms
      {
         ApplicationControlled On
         SecondaryMap  On
         TLSv1 Off
         TLSv1.1 On
         TLSv1.2 On
         TLSv1.3 On
      }
      TTLSCipherParmsRef   ftp_server_ciphers     # Used to customize ciphersuites for the FTP
                                                  # server
   }
TTLSCipherParms ftp_server_ciphers
{
   # Sample ciphers.  Should be customized!
   V3CipherSuites TLS_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_DHE_DSS_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_DHE_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites TLS_DHE_DSS_WITH_AES_128_CBC_SHA
   V3CipherSuites TLS_DHE_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites TLS_AES_128_GCM_SHA256
   V3CipherSuites TLS_AES_256_GCM_SHA384
}

   TTLSRule secure_ftp_server_rule
   {
      LocalPortRange 21                  # This should be set to the port the FTP server is
                                         #  listening on
      Direction Inbound
      TTLSGroupActionRef secure_ftp_server_group
      TTLSEnvironmentActionRef secure_ftp_server_env
   }
```

Tip: You can configure many other TLS settings in AT-TLS policy.

When you define your AT-TLS policy rules for your FTP server, you need to decide on a variety of important TLS settings, such as the acceptable TLS protocol versions and the acceptable cipher suites. If need be, work with your network security administrator to determine the proper values for these settings. For more information about AT-TLS policy statements, see AT-TLS policy statements in z/OS Communications Server: IP Configuration Reference.

9. Decide the level of security for the data connection.

   You can choose to require enciphered data transfers, or to allow the client to decide the level of security for data transfers. The default is to allow the clients to decide the level of security.

This setting is customized by using the SECURE_DATACONN configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

If you want the server to require that data is transferred raw with no cipher algorithm applied to the data and that clients attempting to use ciphers are rejected, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_DATACONN NEVER
```

If you want the client to decide whether data is transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

```
SECURE_DATACONN CLEAR
```

This is the default.

For TLS, the client decides whether data is enciphered or not. If it indicates it should be enciphered, the cipher algorithm is negotiated between the server and the client using TLS protocols. For Kerberos, the client can specify whether data is transferred raw, integrity protected only, or both integrity and privacy protected.

If you want the server to require that data is transferred enciphered and that clients attempting to send raw data are rejected, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_DATACONN PRIVATE
```

For TLS, the cipher algorithm is negotiated between the server and the client using TLS protocols. For Kerberos, the data must be transferred using both integrity and privacy protection. Clients attempting to send data that is only integrity protected are rejected.

10. Decide whether the server requires session reuse when SSL/TLS is used to protect the control and data connections.

    By default, the server is enabled to reuse either of the following SSL session IDs on the data connections within an FTP session:

    - The SSL session ID of the control connection
    - The SSL session ID of a previous data connection

    This setting is customized by using the SECURE_SESSION_REUSE configuration statement.

    - To enable the server to reuse either of the following SSL session IDs on the subsequent data connections within an FTP session, code ALLOWED on the SECURE_SESSION_REUSE statement in the FTP.DATA configuration file of the server:
      - The SSL session ID of the control connection
      - The SSL session ID of a previous data connection

      This is the default.
    - To require the server to reuse the SSL session ID of the control connection on the subsequent data connections within an FTP session, code REQUIRED on the SECURE_SESSION_REUSE statement in the FTP.DATA configuration file of the server.

      This setting might cause data connections and FTP transfers to fail when the remote side does not support reusing the session ID.

    For information about the SECURE_SESSION_REUSE statement, see z/OS Communications Server: IP Configuration Reference.
11. For information about configuring your security product for TLS, see Appendix B, “TLS/SSL security,” on page 1329.

Steps for customizing the FTP server for Kerberos

The FTP server can be enabled to support both TLS and Kerberos.

Before you begin

Some configuration statement settings apply to both TLS and Kerberos and affect the behavior of both.

Decide which RACF ID the service principal is to be associated with, which helps determine whether a keytab file is required. If the service principal is associated with the FTP startup procedure ID, a keytab file is not required. If a keytab file is not required and you do not plan to use one, decide how the FTP startup procedure is to be updated to identify the environment variable (ENVAR) KRB5_SERVER_KEYTAB.

Procedure

Perform the following steps to customize the FTP server for Kerberos:

1. Code the following statement in the server's FTP.DATA configuration file to enable the server for Kerberos:

   ```
   EXTENSIONS AUTH_GSSAPI
   ```

2. Decide whether clients should be required to use the Kerberos protocol.

   The default is to allow the client to decide whether to use Kerberos.

   This setting is customized using the SECURE_FTP configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   To allow the client to decide whether to use Kerberos, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_FTP ALLOWED
   ```

   This is the default setting, and indicates:

   - If the server is enabled for TLS only, clients must either log in using TLS, or with no security mechanism.
   - If the server is enabled for Kerberos only, clients must either log in using Kerberos, or with no security mechanism.
   - If the server is enabled for both TLS and Kerberos, clients can log in using TLS, Kerberos, or with no security mechanism.

   To require that clients log in using Kerberos, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_FTP REQUIRED
   ```

   This setting indicates:

   - If the server is enabled for TLS only, clients must log in using TLS.
   - If the server is enabled for Kerberos only, clients must log in using Kerberos.
   - If the server is enabled for both TLS and Kerberos, clients must log in using either TLS or Kerberos.

3. Decide whether to use the client authentication process to eliminate the client login password prompt so that a client supplies only the login user ID to establish the session.

   The Kerberos principal that is received from the client is used to query the security product (either RACF or another SAF-compliant security product) to determine whether the Kerberos principal maps to a user ID that is known to the system. If the Kerberos principal maps to a user ID, and that user ID matches the user name passed from the client on the USER command, you can eliminate the password prompt.

If the client principal is for the same realm as the FTP server, the principal is correlated to the user ID using the KERBNAME option of the ADDUSER or ALTUSER commands. If the client principal is a cross-realm principal, it is correlated to the user ID using the RDEFINE KERBLINK command.

If you want to require the client to provide a password or password phrase even when the client authentication process does not require it, code the following statement in the server's FTP.DATA configuration file. This is the default.

```
SECURE_PASSWORD_KERBEROS REQUIRED
```

If you want to use the client authentication process to eliminate the client password prompt, code the following statement in the server's FTP.DATA configuration file:

```
SECURE_PASSWORD_KERBEROS OPTIONAL
```

4. Decide the level of security for the data connection.

   You can choose to require enciphered data transfers, or to allow the client to decide the level of security for data transfers. The default is to allow the clients to decide the level of security.

   This setting is customized using the SECURE_DATACONN configuration statement. You should understand that its setting affects both TLS security behavior and Kerberos security behavior.

   If you want the server to require that data is transferred raw with no cipher algorithm applied to the data and that clients attempting to use ciphers are rejected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN NEVER
   ```

   If you want the client to decide whether data is transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN CLEAR
   ```

   This is the default.

   For TLS, the client decides whether data is enciphered or not. If it indicates it should be enciphered, the cipher algorithm is negotiated between the server and the client using TLS protocols. For Kerberos, the client can specify whether data is transferred raw, integrity protected only, or both integrity and privacy protected.

   If you want the server to require that data is transferred both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN PRIVATE
   ```

   For TLS, the cipher algorithm is negotiated between the server and the client using TLS protocols, and clients attempting to send raw data are rejected. For Kerberos, the data must be transferred using both integrity and privacy protection, and clients attempting to send raw data or data that is only integrity protected are rejected.

   If you want the server to require that data is transferred integrity protected only or both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_DATACONN SAFE
   ```

   For TLS, specifying this option is identical to specifying SECURE_DATACONN PRIVATE. For Kerberos, specifying this option indicates the data can be transferred integrity protected only, or both integrity and privacy protected. Clients attempting to send raw data are rejected.

5. Ensure that the RACF user ID under which the FTP client will execute has permission to the ICSF callable services that Kerberos needs. Specifically, if the CSFSERV class is active, ensure that the user ID has read access to the following ICSF SAF resources:

   CSF1TRC, CSF1SKE, CSF1SKD, CSF1TRD, and CSFOWH

6. Decide the level of security for the control connection (that is, for FTP commands and replies).

   You can choose to require enciphered control connection data, or to allow the client to decide the level of security. The default is to allow the clients to decide the level of security.

   This setting is customized using the SECURE_CTRLCONN configuration statement. This setting applies only to Kerberos. For TLS, the control connection is required to be enciphered and this setting has no effect on TLS behavior.

   If you want the client to decide whether control data is transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN CLEAR
   ```

   This is the default.

   The client can specify whether data is transferred raw, integrity protected only, or both integrity and privacy protected.

   If you want the server to require that control data is transferred both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN PRIVATE
   ```

   Clients attempting to send raw data or data that is only integrity protected are rejected.

   If you want the server to require that data is transferred integrity protected only or both integrity and privacy protected, code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN SAFE
   ```

   Clients attempting to send raw data are rejected.
7. Create the service principal against a RACF ID for use with a keytab (see step 7 for using Kerberos with no keytab).
   a) Create a RACF user ID to associate with the FTP service principal.

      ```
      adduser FTP NOPASSWORD DFLTGRP(SYS1) omvs(autouid home('/u/ftp') prog('/bin/
      sh'))
      ```

   b) After the FTP RACF user ID is created, add the Kerberos principal to it.

      ```
      ALTUSER FTP KERB(KERBNAME(ftp/<hostname>))
      ```

   c) To ensure the Kerberos segment was added, use the following command to display the ID.

      ```
      LU FTP NORACF KERB
      ```

      Result:

      ```
      USER=FTP

      KERB INFORMATION
      ----------------
      KERBNAME= ftp/<hostname>
      KEY VERSION= 001
      KEY ENCRYPTION TYPE= DES DES3 DESD
      ```

   d) To add the FTP service principal to the keytab file, take the following actions:
      i) The keytab file is located in the /etc/skrb directory. Switch to that directory using the following command:

         ```
         cd /etc/skrb
         ```

      ii) Use the following command to see what is currently in the keytab file:

```
keytab list
```

If nothing is currently in the keytab file, the following information is returned:

```
Key table: /etc/skrb/krb5.keytab
```

iii) Add the FTP service principle using the following command:

```
keytab add ftp/<hostname>
```

You will be prompted for the principals' password. For this example, that password is FTP. The password must be entered in uppercase. This password was assigned with the RACF ALTUSER command when the FTP service principal was created.

iv) Issue the keytab list command again.

The following information is displayed when the FTP service principal is present:

```
Key table: /etc/skrb/krb5.keytab

Principal: ftp/<hostname>@<realm>
  Key version: 1
  Key type: 56-bit DES
  Entry timestamp: 2005/02/04-16:21:10

Principal: ftp/<hostname>@<realm>
  Key version: 1
  Key type: 56-bit DES using key derivation
  Entry timestamp: 2005/02/04-16:21:10

Principal: ftp/<hostname>@<realm>
  Key version: 1
  Key type: 168-bit DES using key derivation
  Entry timestamp: 2005/02/04-16:21:10
```

8. An alternate way to run without a keytab file is to associate the FTP service principal to the ID under which the FTP started task runs.

   If the ID that the FTP started task runs under is FTPD, issue the following command to create the FTP service principal and have it associated to that ID.

   ```
   ALTUSER FTPD PASSWORD(ftpd) NOEXPIRED KERB(KERBNAME(ftp/<hostname>))
   ```

   Rule: In this setup, you must set the KRB5_SERVER_KEYTAB environment variable. Specify it directly in the FTP startup procedure as follows:

   ```
   //FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
   //      PARM=('POSIX(ON) ALL31(ON)',
   //        'ENVAR("KRB5_SERVER_KEYTAB=1")/&PARMS')
   ```

   Another way of specifying the environment variable directly in the startup procedure is to specify a file where the environment variables are listed.

   ```
   //FTPD   EXEC PGM=&MODULE,REGION=4096K,TIME=NOLIMIT,
   //      PARM=('POSIX(ON) ALL31(ON)',
   //        'ENVAR("_CEE_ENVFILE=/etc/ftp.envvars")/&PARMS')
   ```

   Then, within the /etc/ftp.envvars file, add the following entry:

   ```
   KRB5_SERVER_KEYTAB=1
   ```

9. Ensure that the FTP RACF user ID has permission to the ICSF callable services that Kerberos needs. Specifically, if the CSFSERV class is active, ensure that the user ID has read access to the following ICSF SAF resources:

   CSF1TRC, CSF1SKE, CSF1SKD, CSF1TRD, and CSFOWH

Results

You know you are done when a client is able to successfully log in to the FTP server using Kerberos. An example of the login is as follows:

1. Obtain the Kerberos credentials by issuing the following command:

   ```
   kinit joe
   ```

2. You will be prompted for the password. Enter it.
3. Issue the **ftp** command:

   ```
   ftp <hostname>
   ```

   You should see the following information:

   ```
   Using /u/JOE/ftp.data for local site configuration parameters.
   IBM FTP CS V1R9
   FTP: using TCPIP
   Connecting to: <hostname><ip address> port:  <port number>.
   220-FTPD1 IBM FTP CS V1R9 at <hostname>, 21:51:51 on 2007-04-04.
   220 Connection will close if idle for more than 5 minutes.
   >>> AUTH GSSAPI
   334 Using authentication mechanism GSSAPI
   >>> ADAT
   235 ADAT=YGgGCSqGSIb3EgECAgIAb1kwV6ADAgEFoQMCAQ+iSzBJoAMC7moS==
   Authentication negotiation succeeded
   NAME (<hostname>:USER):
   JOE
   >>> USER JOE
   331 Send password please.
   PASSWORD:

   >>> PASS
   230 JOE is logged on.  Working directory is "JOE".
   Command:
   ```

   Tip: The password prompt is skipped if the server is configured with SECURE_PASSWORD_KERBEROS OPTIONAL and the client's Kerberos ticket principal name matches the logon user ID.

Steps for customizing the FTP client for TLS

You can customize the FTP client for TLS, but a better way to implement TLS security is by using AT-TLS.

Before you begin

Understand the following information:

- The FTP client can be enabled to use either TLS or Kerberos, but not both at the same time.
- To support TLS, the FTP server always provides server certificate authentication to all the clients to validate that the server is what it says it is. Therefore, a client key ring database is required to contain at least the certificate for the CA that signed the server certificate (or the server certificate if the server certificate is self-signed). For more information about key ring databases, see Appendix B, “TLS/SSL security,” on page 1329.
- The FTP client can implement TLS security by itself, or the FTP client can be configured to use Application Transparent Transport Layer Security (AT-TLS) as a controlling application. For more information on AT-TLS, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

  Guideline: Using AT-TLS is the best way to implement TLS security. With AT-TLS, for example, you can have the following implementation:

  – Specify the label of the certificate to be used for authentication instead of using the default certificate
  – Support SSL Session Key Refresh
  – Support SSL Sysplex Session ID Caching
  – Trace decrypted SSL data for FTP in a data trace

– Receive more detailed diagnostic messages in syslogd

Requirement: AT-TLS requires Policy Agent to be configured, and the TCP/IP stack to be enabled for AT-TLS. To configure AT-TLS, see “Configuring the client systems” on page 1128.

Procedure

Perform the following steps to customize the FTP client for TLS:

1. Decide what level of RFC 4217, *On Securing FTP with TLS*, that you want the client to support.
   - To have the client support *On Securing FTP with TLS* at the Internet draft level, code the following statement in the client's FTP.DATA configuration file:

     ```
     TLSRFCLEVEL DRAFT
     ```

     This is the default. The z/OS FTP client has supported TLS security at this level since V1R2. Code this statement in FTP.DATA to maintain this level of support.
   - To have the client support *On Securing FTP with TLS* at the RFC 4217 level, code the following statement in the client's FTP.DATA configuration file:

     ```
     TLSRFCLEVEL RFC4217
     ```

     The RFC *On Securing FTP with TLS* was published as RFC 4217 in October, 2005. The RFC differs from the Internet draft in its description of the AUTH, CCC, and REIN commands. This has implications for client subcommands such as AUTH and CCC. Generally, RFC 4217 is less restrictive than the Internet draft. For more information, see RFC 4217. For more information on RFC 4217 and using security mechanisms, see z/OS Communications Server: IP User's Guide and Commands.
2. Code the following statement in the client's FTP.DATA configuration file to enable the client for TLS:

   ```
   SECURE_MECHANISM TLS
   ```
3. Decide what level of authentication you will use for TLS sessions:
   - Server authentication only
   - Client authentication level 1
   - Client authentication level 2
   - Client authentication level 3

   For more information about server authentication and client authentication, see “Secure Socket Layer overview” on page 1329.
4. Use a CERTAUTH virtual key ring, or create a client key ring database and add the certificates that you need to that database.

   If you are using server authentication only and the FTP server certificate is signed by a certificate authority (CA), the FTP client can use a CERTAUTH virtual key ring and you do not need to create a client key ring database. To use a CERTAUTH virtual key ring, use the key ring name ∗AUTH∗/∗ .

   If you cannot use a virtual key ring, create the client key ring database and add the certificates that you need to that database.

   Every TLS session handshake includes server authentication. If a server certificate is self-signed, you must import that certificate to the key ring database of any client that will log in using TLS. If the server certificate is signed by a CA, the CA certificate used to sign the server certificate (rather than the server certificate itself) needs to be in the client key ring database. For more information, see “Server authentication ” on page 1329.

   If you are using client authentication, you must add a certificate for the client to the client key ring database.

   If you are using client authentication and self-signed client certificates, you must add a certificate for the client to the server key ring database. If a client certificate is signed by a CA, the CA certificate

used to sign the client certificate needs to be in the server key ring database, rather than the client certificate.

For information about the client certificates you must create, see "Client authentication" on page 1330.

5. Decide whether FTP will implement TLS security or AT-TLS will implement TLS security.

   The default is to have FTP implement TLS security. This setting is customized using the TLSMECHANISM configuration statement.

   - To configure the FTP client to use AT-TLS for TLS security, code the following statement in FTP.DATA:

     ```
     TLSMECHANISM ATTLS
     ```

   - To configure the FTP client to implement TLS security by itself, code the following statement in FTP.DATA:

     ```
     TLSMECHANISM FTP
     ```

     This is the default setting.

6. If using TLSMECHANISM FTP, you must configure the FTP client with the name of the key ring database.

   Code the following statement in FTP.DATA:

   ```
   KEYRING client-keyring-database
   ```

   For information about the KEYRING statement, see z/OS Communications Server: IP Configuration Reference.

   Tip: The key ring name can optionally be qualified with the user ID that owns the key ring (*userID/client-keyring-database*). If the FTP.DATA data set will be shared by multiple z/OS users, then omitting *userID* ensures that each user will use their own instance of the keyring. If the FTP.DATA data set is used by a single z/OS user, then including the *userID* can help simplify diagnostic efforts.

7. If you specified TLSMECHANISM ATTLS, configure the AT-TLS policy for the FTP client.

   To configure AT-TLS, see "Configuring the client systems" on page 1128.

   Requirements:

   - The FTP server and client are controlling applications. For more information about controlling applications, see "Advanced application considerations" on page 1144.

     Code a TTLSEnvironmentAdvancedParms statement with the ApplicationControlled and SecondaryMap parameters; both parameters should specify the value On. The ApplicatonControlled parameter allows FTP to start and stop TLS security on a connection. The SecondaryMap parameter enables active or passive data connections to use the AT-TLS policy that is used for the control connection. You do not need to code any additional TTLSRule statements for the data connections.
   - The FTP client requires the HandshakeRole parameter with the value Client to be coded on the TTLSEnvironmentAction statement.
   - The TTLSRule statement for the FTP client requires the Direction parameter with the value Outbound.

   A sample Policy Agent AT-TLS configuration showing the required policy configuration statements for AT-TLS is as follows:

   ```
   TTLSGroupAction  secure_ftp_client_group
   {
      TTLSEnabled On
   }
   TTLSEnvironmentAction secure_ftp_client_env
   {
      TTLSKeyringParms
   ```

```
      {       # The key ring value can optionally be qualified with
the user ID that        # owns the key ring (userID/client-keyring-
database).                          # If the AT-TLS rule will be shared by
multiple z/OS users, then             # omitting userID ensures that each user
will use their own instance of       # the keyring. If the AT-TLS rule is used
by a single z/OS user, then         # including the userID can help simplify
diagnostic efforts.
         Keyring client-keyring-database
      }
      HandshakeRole Client
      TTLSEnvironmentAdvancedParms
      {
         ApplicationControlled On
         SecondaryMap  On
         TLSv1 Off
         TLSv1.1 On
         TLSv1.2 On
         TLSv1.3 On
      }
      TTLSCipherParmsRef   ftp_client_ciphers   # Used to customize ciphersuites
   }
TTLSCipherParms ftp_client_ciphers
{
   # Sample ciphers.  Should be customized!
   V3CipherSuites TLS_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_DHE_DSS_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_DHE_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites TLS_DHE_DSS_WITH_AES_128_CBC_SHA
   V3CipherSuites TLS_DHE_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites TLS_AES_128_GCM_SHA256
   V3CipherSuites TLS_AES_256_GCM_SHA384
}

   TTLSRule secure_ftp_client_rule
   {
      RemotePortRange 21       # This should be set to the port the FTP server is
                               # listening on
      Direction Outbound
      TTLSGroupActionRef secure_ftp_client_group
      TTLSEnvironmentActionRef secure_ftp_client_env
   }
```

Tip: You can enable additional security settings with AT-TLS, such as LDAP servers and handshake timeout values. The sample configuration is only the minimum required to allow the FTP client to use AT-TLS. You can add additional configuration statements.

8. Decide which cipher algorithms the client should use to encipher data transfers and control information.

   FTP and AT-TLS support TLS through the system SSL cryptographic services base element of z/OS. System SSL supports multiple cipher algorithms that provide both encryption and data authentication (that is, data integrity). Encryption scrambles the data so it is transferred confidentially and cannot be interpreted without a special key. Data authentication algorithms ensure that the data was not modified during transfer. Some of the supplied cipher algorithms provide only data authentication, and some provide both encryption and authentication. Be aware that the actual cipher algorithm used for the session is determined by a negotiation between the server and client. For example, if you configure an FTP client to use the Triple DES encryption, SHA authentication algorithm, but the server does not support that cipher algorithm, Triple DES encryption, SHA authentication will not be used for sessions between the client and that server.

   If using TLSMECHANISM FTP, select which cipher algorithms you prefer to use by coding a CIPHERSUITE configuration statement in the FTP.DATA file for each cipher algorithm the client can use. For a list of the cipher algorithms you can specify on the CIPHERSUITE statement, see z/OS Communications Server: IP Configuration Reference.

   If you specify TLSMECHANISM ATTLS, select which cipher algorithms you want to use by coding a TTLSCipherParms configuration statement to specify the cipher algorithms the client can use. For a list of the cipher algorithms you can specify with the TTLSCipherParms statement, see z/OS

Communications Server: IP Configuration Reference. List the ciphers in the order of preference, your most preferred cipher algorithm first. The cipher algorithm is negotiated with the server on behalf of the client using the same order of preference as is indicated by the order of the TTLSCipherParms statement.

Restrictions:

- Only RSA key exchange is supported.
- The following algorithms are subject to export regulations and might not be available to your system:
  - Triple DES encryption, SHA authentication
  - RC4 (128-bit) encryption, SHA authentication
  - RC4 (128-bit) encryption, MD5 authentication
  - AES (128-bit and 256-bit) encryption, SHA authentication

9. Decide whether the client should be required to use the TLS protocol.

   If the FTP server does not support TLS, you can choose to allow the client to log in without using the TLS security, or require the client to use a secure session, thus failing the login. The default is to not require the client to use TLS. This setting is customized using the SECURE_FTP configuration statement.

   To have the client log in using the TLS protocol when the server supports TLS, and log in without TLS when the server does not support TLS, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_FTP ALLOWED
   ```

   This is the default.

   To have the client log in using the TLS protocol, but close the server connection and prevent logging in when the server does not support TLS, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_FTP REQUIRED
   ```

10. Decide the level of security for the data connection.

    You can choose to require enciphered data transfers, or to allow the FTP user to decide the level of security for data transfers. The default is to not encipher the data, but allow the data to be enciphered at the server's request or at the FTP user's request during the FTP session.

    Note that the level of security for data connections is determined by both the SECURE_DATACONN statement in FTP.DATA and by subcommands an FTP user might issue during an FTP session.

    The following subcommands can be issued by the user:

    clear
    Resets the security level so that data is transferred raw.

    private
    Resets the security level so that data is transferred enciphered. The cipher algorithm is negotiated between the server and the client using the TLS protocol negotiation.

    If you want the client to transfer data raw with no cipher algorithm applied to the data, code the following statement in the client's FTP.DATA configuration file:

    ```
    SECURE_DATACONN NEVER
    ```

    To indicate the data can be transferred raw or enciphered, code the following statement in the client's FTP.DATA configuration file:

    ```
    SECURE_DATACONN CLEAR
    ```

This is the default.

By default, data is transferred raw. However, the user can issue the `private` subcommand during the FTP session to change the data connection security level, so that data is transferred enciphered. The user can also issue the `clear` subcommand to reset the data connection security level back, so that data is transferred raw again. For TLS, if the `private` subcommand is issued, the cipher algorithm is negotiated between the server and the client using TLS protocols.

If you want to require that data is transferred enciphered, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_DATACONN PRIVATE
```

For TLS, the cipher algorithm is negotiated between the server and the client using TLS protocols.

11. Decide whether the client requires session reuse when SSL/TLS is used to protect the control and data connections.

    By default, the setting is not enabled for the client to reuse either of the following SSL session IDs on subsequent data connections within an FTP session:

    - The SSL session ID of the control connection
    - The SSL session ID of a previous data connection

    This setting is customized by using the SECURE_SESSION_REUSE configuration statement.

    - If you do not require the client to reuse either of the following SSL session IDs on subsequent data connections within an FTP session, code NONE on the SECURE_SESSION_REUSE statement in the FTP.DATA configuration file of the client:
      - The SSL session ID of the control connection
      - The SSL session ID of a previous data connection

      This is the default.
    - To enable the client to reuse either of the following SSL session IDs on subsequent data connections within an FTP session, code ALLOWED on the SECURE_SESSION_REUSE statement in the FTP.DATA configuration file of the client:
      - The SSL session ID of the control connection
      - The SSL session ID of a previous data connection
    - To require the client to reuse the SSL session ID of the control connection on the subsequent data connections within an FTP session, code the SECURE_SESSION_REUSE REQUIRED statement in the FTP.DATA configuration file of the client.

      This setting might cause data connections and FTP transfers to fail when the remote side does not support reusing the session ID.

    For information about the SECURE_SESSION_REUSE statement, see z/OS Communications Server: IP Configuration Reference.

Steps for customizing the FTP client for Kerberos

The FTP client can be enabled to use either TLS or Kerberos, but not both at the same time.

Procedure

Perform the following steps to customize the FTP client for Kerberos:

1. Code the following statement in the client's FTP.DATA configuration file to enable the client for Kerberos:

   ```
   SECURE_MECHANISM GSSAPI
   ```

2. Decide whether the client should be required to use the Kerberos protocol.

If the FTP server does not support Kerberos, you can choose to allow the client to log in without using Kerberos security, or require the client to use a secure session, thus failing the login. The default is to not require the client to use Kerberos. This setting is customized using the SECURE_FTP configuration statement.

To have the client log in using the Kerberos protocol, but if the server does not support Kerberos allow the client to complete the login without using it, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_FTP ALLOWED
```

This is the default.

To have the client log in using the Kerberos protocol, but if the server does not support Kerberos have the login fail and not allow the client to log in, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_FTP REQUIRED
```

3. Decide the level of security for the data connection.

   You can choose to require enciphered data transfers, or to allow the FTP user to decide the level of security for data transfers. The default is to not encipher the data, but allow the data to be enciphered at the server's request or at the FTP user's request during the FTP session.

   Note that the level of security for data connections is determined by both the SECURE_DATACONN statement in FTP.DATA and by subcommands an FTP user might issue during an FTP session.

   The following subcommands can be issued by the user:

   clear
   Resets the security level so that data is transferred raw.

   private
   Resets the security level so that data is transferred enciphered. If the client is using the Kerberos security mechanism, the data is transferred both integrity protected and privacy protected. If the client is using the TLS security mechanism, the cipher algorithm is negotiated between the server and the client using the TLS protocol negotiation.

   safe
   Resets the security level so that data is transferred integrity protected only.

   If you want the client to transfer data raw with no cipher algorithm applied to the data, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN NEVER
   ```

   To indicate the data can be transferred raw or enciphered, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN CLEAR
   ```

   This is the default.

   By default, data is transferred raw. However, the user can issue the `private` subcommand during the FTP session to change the data connection security level, so that data is transferred both integrity and privacy protected. The user can also issue the `safe` subcommand to change the data connection security level so that data is transferred integrity protected only, or the `clear` subcommand to reset the data connection security level back so that data is transferred raw again.

   If you want to require that data is transferred both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_DATACONN PRIVATE
   ```

If you want to require that data is transferred integrity protected only, or both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

```
SECURE_DATACONN SAFE
```

By default, data is transferred integrity protected only. However, the user can issue the `private` subcommand during the FTP session to change the data connection security level so that data is transferred both integrity and privacy protected. The user can also issue the `safe` subcommand to reset the data connection security level back, so that data is transferred integrity protected only.

4. Decide the level of security for the control connection (that is, for FTP commands and replies).

   You can choose to require enciphered data, or to allow the FTP user to decide the level of security. The default is to not encipher the data, but allow the data to be enciphered at the server's request or at the FTP user's request during the FTP session.

   Note that the level of security for data connections is determined by both the SECURE_CTRLCONN statement in FTP.DATA and by subcommands an FTP user might issue during an FTP session.

   The following subcommands can be issued by the user:

   cprotect clear
   Resets the security level so that data is transferred raw.

   cprotect private
   Resets the security level so that data is transferred both integrity protected and privacy protected.

   cprotect safe
   Resets the security level so that data is transferred integrity protected only.

   To indicate the data can be transferred raw or enciphered, you can code the following statement in the server's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN CLEAR
   ```

   This is the default.

   By default, data is transferred raw. However, the user can issue the `cprotect private` subcommand during the FTP session to change the security level so that data is transferred both integrity and privacy protected. The user can also issue the `cprotect safe` subcommand to change the security level so that data is transferred integrity protected only, and the `cprotect clear` subcommand to reset the security level back so that data is transferred raw again.

   If you want to require that data is transferred both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN PRIVATE
   ```

   If you want to require that data is transferred integrity protected only, or both integrity and privacy protected, code the following statement in the client's FTP.DATA configuration file:

   ```
   SECURE_CTRLCONN SAFE
   ```

   By default, data is transferred integrity protected only. However, the user can issue the `cprotect private` subcommand during the FTP session to change the data connection security level so that data is transferred both integrity and privacy protected. The user can also issue the `cprotect safe` subcommand to reset the data connection security level, so that data is transferred integrity protected only.

Port 990

The use of port 990 to implicitly protect FTP sessions was included in the early drafts of the IETF documents that describe how to use TLS with FTP, but has been removed from later drafts and from RFC 4217. For more information, see Information APAR II13516.

Port 990 is known as the protected port, or the TLSPORT. You can disable implicit security for port 990, or reassign the protected port, by coding the TLSPORT statement in the server's FTP.DATA configuration file.

Rule: If you start the FTP server on the protected port, you should code a SECUREIMPLICITZOS statement in the server's FTP.DATA file to specify when the server should expect the client to negotiate TLS security.

The FTP server can provide explicit TLS security on a different port by specifying the following definitions in FTP.DATA:

```
EXTENSIONS AUTH_TLS
SECURE_FTP REQUIRED
SECURE_CTRLCONN PRIVATE
SECURE_DATACONN PRIVATE
```

Steps for migrating the FTP client to use AT-TLS

Application Transparent Transport Layer Security (AT-TLS) is the best way to implement TLS security for the FTP server and client. AT-TLS provides additional functionality and performance for TLS secured connections.

Procedure

Perform the following steps to migrate from an existing configuration using TLS security for the FTP client to a configuration using AT-TLS:

1. Configure AT-TLS and Policy Agent.

   For details about AT-TLS setup, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121. For Policy Agent setup and AT-TLS policy statements, see z/OS Communications Server: IP Configuration Reference.

   Requirements:

   - The FTP client is a controlling application. For more information about controlling applications, see “Advanced application considerations” on page 1144.

     Code a TTLSEnvironmentAdvancedParms statement with the ApplicationControlled and SecondaryMap parameters; both parameters should specify the value On. The ApplicatonControlled parameter allows FTP to start and stop TLS security on a connection. The SecondaryMap parameter enables active or passive data connections to use the AT-TLS policy that is used for the control connection. You do not need to code any additional TTLSRule statements for the data connections.
   - The FTP client requires the HandshakeRole parameter with the value Client to be coded on the TTLSEnvironmentAction statement.
   - The TTLSRule statement for the FTP client requires the Direction parameter with the value Outbound.

   Guideline: The FTP client does not support SSLv2 when using TLSMECHANISM FTP. By default, AT-TLS does not enable SSLv2. SSLv2 should not be enabled in AT-TLS unless explicitly required by a remote system. If SSLv2 is required by a remote system, use a specific TTLSRule statement for the remote system that points to a TTLSConnectionAction statement enabling SSLv2.
2. Configure the FTP client to use AT-TLS by coding TLSMECHANISM ATTLS in FTP.DATA.
3. If TLSRFCLEVEL CCCNONOTIFY is configured in FTP.DATA, update TLSRFCLEVEL to use a value that is supported with AT-TLS.
4. Use Table 32 on page 704 to migrate the existing FTP client configuration to AT-TLS.

   Remove the statements from FTP.DATA and code the AT-TLS equivalent statement.

Table 32. Migrating existing FTP client configuration		
FTP.DATA statement	**AT-TLS equivalent statement**	**AT-TLS policy statement**
KEYRING	Keyring	TTLSKeyRingParms -> TTLSEnvironmentAction
CIPHERSUITE	V3CipherSuites	TTLSCipherParms -> TTLSEnvironmentAction or TTLSConnectionAction
TLSTIMEOUT	GSK_V3_SESSION_TIMEOUT	TTLSGskAdvancedParms -> TTLSEnvironmentAction
SSLV3	SSLv3	TTLSEnvironmentAdvancedParms-> TTLSEnvironmentAction Or TTLSConnectionAdvancedParms-> TTLSConnectionAction

5. Use Table 33 on page 704 to migrate existing ciphers coded on CIPHERSUITE statements in FTP.DATA to AT-TLS TTLSCipherParms statements.

Table 33. Migrating existing ciphers		
CIPHERSUITE cipher	**V3CipherSuites cipher**	**Hexadecimal value**
SSL_DES_SHA	TLS_RSA_WITH_DES_CBC_SHA	09
SSL_3DES_SHA	TLS_RSA_WITH_3DES_EDE_CBC_SHA	0A
SSL_NULL_MD5	TLS_RSA_WITH_NULL_MD5	01
SSL_NULL_SHA	TLS_RSA_WITH_NULL_SHA	02
SSL_RC2_MD5_EX	TLS_RSA_EXPORT_WITH_RC2_CBC_40_MD5	06
SSL_RC4_MD5	TLS_RSA_WITH_RC4_128_MD5	04
SSL_RC4_MD5_EX	TLS_RSA_EXPORT_WITH_RC4_40_MD5	03
SSL_AES_128_SHA	TLS_RSA_WITH_AES_128_CBC_SHA	2F
SSL_AES_256_SHA	TLS_RSA_WITH_AES_256_CBC_SHA	35

For example, for an FTP.DATA file that contains the following statements:

```
CIPHERSUITE SSL_AES_256_SHA
CIPHERSUITE SSL_3DES_SHA
CIPHERSUITE SSL_NUL_SHA
```

The equivalent TTLSCipherParms statement:

```
TTLSCipherParms
{
   V3CipherSuites TLS_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites TLS_RSA_WITH_3DES_EDE_CBC_SHA
   V3CipherSuites TLS_RSA_WITH_NULL_SHA
}
```

6. AT-TLS supports more secure TLS versions and ciphers. Consider enabling TLSv1.2 or later on the TTLSEnvironmentAdvancedParms or TTLSConnectionAdvancedParms statement. Consider enabling support for additional ciphers on the TTLSCipherParms statement.

Traversing firewalls with SSL/TLS secure FTP

This topic describes functions in FTP that enable you to use FTP sessions that are secured by SSL/TLS using both network address translation (NAT) and filtering firewalls.

FTP requires the following TCP connections to transfer a file:

- Control connection
- Data connection

The control connection is established from the FTP client to the FTP server (default port 21). The data connection is established either from the FTP client to the FTP server, or from the FTP server to the FTP client; the direction is based on whether the client selects active mode or passive mode FTP.

- Active mode

  With active mode FTP, the data connection is established from the FTP server to the FTP client, which is the opposite direction of the control connection. The active mode data connection is established from a well-known port on the server host (default port 20) to an ephemeral port on the client host.

- Passive mode

  With passive mode FTP, the data connection is established from the FTP client to the FTP server, which is the same direction as the control connection. The data connection is established from an ephemeral port on the server host to an ephemeral port on the client host.

  Passive mode is also referred to as firewall-friendly FTP. An intranet FTP client connecting to an Internet FTP server can establish connections outbound through the company firewall, but not inbound through the firewall. With passive mode, both the control and data connections are established outbound through the firewall to the Internet.

The FTP client user decides which mode to use. Active mode is the default, but the user can usually change to passive mode. The z/OS FTP client user can switch between active and passive modes by issuing the LOCSITE subcommand with the NOFWFRIENDLY and FWFRIENDLY parameters.

Both active mode and passive mode FTP require the exchange of IP address and port information over the control connection. For active mode, the FTP client sends a PORT command specifying the IP address and port number to which the server must connect to establish the data connection. For passive mode, the FTP client sends a PASV command to the server, and the server replies with the IP address and port number to which the client should connect to establish the data connection.

Firewalls are often aware of FTP; they monitor the exchanges over the FTP control connection to learn the IP address and port number to which the data connection is to be established. NAT firewalls change the IP addresses on the PORT command or in the PASV reply. Filtering firewalls install dynamic filters based on the IP addresses and port information to enable the data connection to be established.

When you use SSL/TLS for FTP, the control connection is typically encrypted, so firewalls between the FTP client and server cannot see the data that is exchanged on the PORT command and the PASV reply. The firewalls cannot perform NAT successfully and they cannot install dynamic filters for the data connection, so the result is that your data connection very likely fails.

z/OS FTP includes the following support for functions that are specifically aimed at enabling FTP sessions through such firewalls:

- Extended passive mode (EPSV)

  Extended passive mode works very much like passive mode. Instead of sending a PASV command to the server, the client sends an EPSV command to the server. The server EPSV reply includes only a port number. The client always uses the same IP address for the data connection that it used for the control connection. A z/OS FTP client user switches to extended passive mode for IPv4 connections by issuing the LOCSITE subcommand with the EPSV4 and FWFRIENDLY parameters. These options can also be configured in the z/OS FTP client FTP.DATA file.

  EPSV allows sessions secured by SSL/TLS through NAT firewalls, but EPSV alone does not allow FTP sessions that are secured by SSL/TLS through firewalls that also implement dynamic filters.

- The PASSIVEIGNOREADDR configuration option

  This z/OS FTP client option directs the z/OS FTP client to ignore the IP address in the PASV reply and use only the port number when FTP is in passive mode. The client uses the same IP address that it used to log in to the FTP server for the data connection. A z/OS FTP client user enables this support by issuing a LOCSITE subcommand with the PASSIVEIGNOREADDR option. You can also configure this option in the z/OS FTP client FTP.DATA file.

  The PASSIVEIGNOREADDR configuration enables sessions secured by SSL/TLS through NAT firewalls in the same way that extended passive mode enables them, subject to the same limitations. You can use the PASSIVEIGNOREADDR option when the server does not support the EPSV command.
- The PASSIVEDATAPORTS statement in FTP.DATA

  This z/OS FTP server option enables you to define a range of port numbers that the z/OS FTP server can use in PASV and EPSV replies for passive mode data connections. If the PASSIVEDATAPORTS statement on the z/OS FTP server is used in combination with EPSV from the FTP client, the two techniques together allow FTP sessions secured with SSL/TLS through NAT firewalls that also implement static IP filters, assuming that the firewall administrators add static filter rules that allow FTP data connections access to one or more of the ports in the PASSIVEDATAPORTS range.
- The clear command channel (CCC) command

  The CCC command can be used on a control connection secured by SSL/TLS to disable SSL/TLS security. The control connection starts out as secured by SSL/TLS, and stays that way until a user ID and password or password phrase have been exchanged with the server. At that point, the FTP client can send a CCC command to the server, which disables SSL/TLS for the control connection and enables PORT commands and replies to PASV and EPSV commands to flow in the clear on the control connection. When these exchanges occur in the clear, firewalls between the FTP client and FTP server can perform NAT processing and dynamic filter processing as if the connection is not secured with SSL/TLS. The CCC command does not turn off security for the data connection.

  To enable the z/OS FTP server to accept the CCC command, configure TLSRFCLEVEL RFC4217 at the FTP server.

  To enable the z/OS FTP client to support the CCC command, configure TLSRFCLEVEL RFC4217 (or, if TLSMECHANISM FTP is specified,TLSRFCLEVEL CCCNONOTIFY) at the FTP client, matching the client TLSRFCLEVEL value to the server TLSRFCLEVEL value. You can use the LOCSITE subcommand to change the TLSRFCLEVEL value. When TLSRFCLEVEL RFC4217 (or TLSRFCLEVEL CCCNONOTIFY, when using TLSMECHANISM FTP) is configured at the z/OS FTP client, use the CCC subcommand after logging in to the FTP server to send a CCC command to the FTP server.

The support that you use depends on your network topology. The following scenarios are a few selected scenarios to consider for making sure that FTP sessions secured by SSL/TLS can get through your network. The scenarios assume that z/OS is at least one of the endpoints of the secure FTP session. The partner endpoint can be z/OS or any secure FTP product on the market that supports the same RFC levels as z/OS (primarily RFC 4217).

- Your firewall performs NAT only (minimal or no filtering), your FTP client is in a private network behind the NAT firewall, and the FTP server is in a public network such as the Internet, as shown in Figure 94 on page 706.

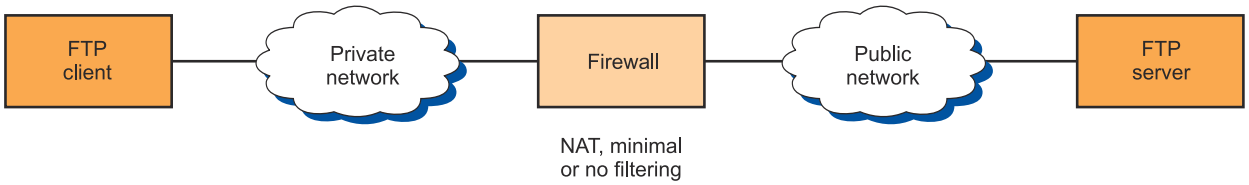


Figure 94. SSL/TLS-secured FTP session scenario 1

Normal passive mode (PASV) usually works in such a scenario. Extended passive mode (EPSV) also works, but is generally not required.

- Your firewalls perform NAT only (minimal or no filtering), your FTP client is in one private network, your FTP server is in another private network, and you have two NAT firewalls between the client and server networks that are connected over a public network, as shown in Figure 95 on page 707.

Figure 95. SSL/TLS-secured FTP session scenario 2

If your partner's secure FTP product supports extended passive mode, use extended passive mode (EPSV) from the FTP client. If the FTP client is the z/OS FTP client and the partner's secure FTP server product does not support EPSV, configure the PASSIVEIGNOREADDR option at your z/OS FTP client to simulate EPSV processing.

- Your firewall performs NAT and static filtering (predefined filter rules). Your FTP client is in a private network behind the NAT firewall, with a z/OS FTP server that is in a public network such as the Internet, as shown in Figure 96 on page 707.

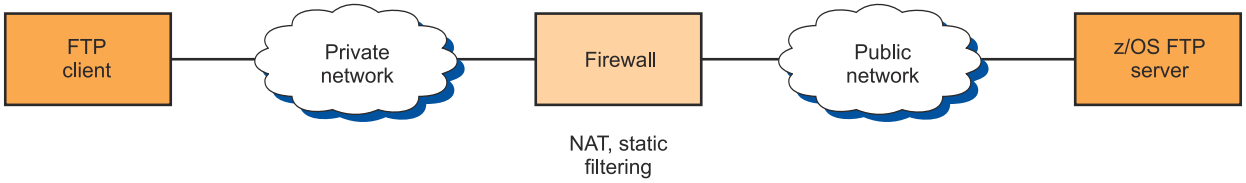


Figure 96. SSL/TLS-secured FTP session scenario 3

Use the PASSIVEDATAPORTS statement in the z/OS FTP server's FTP.DATA file to predefine a range of port numbers that the z/OS FTP server can use for data connections. Your firewall administrator needs to add static filter rules for the passive data port range. Normal passive mode (PASV) usually works in such a scenario, but extended passive mode (EPSV) can also be used if supported by the FTP client.

- Your firewalls perform NAT and static filtering (predefined filter rules). Your FTP client is in one private network, your z/OS FTP server is in another private network, and you have two NAT firewalls between the client and server networks that are connected over a public network, as shown in Figure 97 on page 707.

Figure 97. SSL/TLS-secured FTP session scenario 4

Use the PASSIVEDATAPORTS statement in the z/OS FTP server's FTP.DATA file to predefine a range of port numbers that the z/OS FTP server can use for data connections. Your firewall administrator needs to add static filter rules for the passive data port range. In this case, you must use extended passive mode. If the FTP client does not support extended passive mode, this scenario is not likely to work.

- Your firewalls perform dynamic filtering (with or without NAT) and your partner's secure FTP product supports the CCC command, as shown in Figure 98 on page 707.

Figure 98. SSL/TLS-secured FTP session scenario 5

Use the CCC command from the FTP client. This scenario is not likely to work without CCC command support.

- You do not know what your firewalls do and your partner's secure FTP product does support the CCC command, as shown in Figure 99 on page 708.

Figure 99. SSL/TLS-secured FTP session scenario 6

Use the CCC command from the FTP client. This scenario is not likely to work without CCC command support.

Firewalls reject FTP sessions secured by SSL/TLS in the following additional scenarios:

- Some firewalls are known to apply various validity checks on the FTP control connection data stream. One known check verifies that all interactions on the FTP control connection are terminated with an ASCII newline (NL) character. Most of these checks fail when the control connection is secured with SSL/TLS, because the data is encrypted. If you use the information in this topic and still run into problems establishing FTP sessions secured by SSL/TLS through firewalls, verify with your firewall administrators whether your firewalls implement such validity checks on the FTP control connection, and consider disabling those validity checks.
- Some firewalls are known to disable active mode data connections by default, and block all active mode data connections. Use passive mode or extended passive mode FTP instead.
- Many firewalls monitor activity on TCP connections and terminate connections that are idle for a certain period of time. During a large data transfer over an FTP data connection, the FTP control connection is idle. To avoid having firewalls terminate idle FTP connections, consider coding the FTPKEEPALIVE statement in the z/OS FTP.DATA file for the client or the server. For more information about the FTPKEEPALIVE statement, see “Configuring PROFILE.TCPIP for FTP” on page 664.

Db2 and JES

The following statements are used when FTP interfaces with Db2 and JES, respectively. For more information, see z/OS Communications Server: IP Configuration Reference and the optional steps in this information.

- DB2®
- DB2PLAN
- JESGETBYDSN
- JESINTERFACELEVEL
- JESLRECL
- JESPUTGETTO
- JESRECFM
- SPREAD and SQLCOL

Configuring the optional FTP user exits

This information describes exit routines that you can code and install. For information about coding user exits and for a list of sample user exits, see z/OS Communications Server: IP Configuration Reference.

Tip: You can find the sample user exit routines in SEZAINST.

The FTPSMFEX user exit (for the FTP server)

Note the FTP server SMF user exit is called before an SMF type 118 record that contains information about an FTP server session is written to the SYS1.MANx data set. The user exit allows site specific modifications to the record and controls whether the record is written to the SYS1.MANx data set.

Note that the exit is called only for type 118 records. SMF type 119 FTP records must use the system-wide SMF user exits (IEFU83, IEFU84, and IEFU85) to obtain this same functionality. For information on these SMF user exits, see z/OS MVS System Management Facilities (SMF).

The FTCHKIP user exit (for the FTP server)

The FTCHKIP user exit is called when a user attempts to log in to the FTP server or when a user issues the OPEN subcommand to establish a new connection. The following information is passed to the exit:

- Client IP address *
- Client port number *
- Server IP address *
- Server port number *
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier

Note: Fields above marked with an asterisk (*) are valid only for IPv4 addresses, including IPv4-mapped IPv6 addresses.

An installation can use this exit to determine if a particular IP address or port number is allowed to access the FTP site. If the connection is denied by the user exit, the following message is sent to the user:

```
421 User Exit rejects open for connection
```

The FTCHKPWD user exit (for the FTP server)

The FTCHKPWD user exit is called immediately after the user enters the password, password phrase, or email address while logging in to the FTP server. The following information is passed to the exit:

- The user ID
- The user password

  This field will be set to an asterisk (*) if an email address is entered instead of a password.

  This field will be set to the first eight characters of the password phrase if the user provides a password phrase instead of a password. The entire password phrase is passed to this exit as another parameter.
- A userdata buffer

  If an email address is entered to log in, the userdata buffer contains the email address.
- The number of incorrect passwords or password phrases entered during this session
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier
- The user password or password phrase used to log in to the FTP server

The exit can be used to restrict access to a site based on user ID, password or password phrase, number of bad passwords or password phrases entered, or anything in the socket address information for the client or server. If the login is denied by the user exit, the following reply is sent to the user:

```
530 PASS command failed
```

Result: If you coded ACCESSERRORMSGS TRUE in FTP.DATA, an additional 530 reply with information about why the PASS command failed might precede the reply above.

The FTCHKCMD user exit (for the FTP server)

The FTCHKCMD user exit is called whenever the server receives an FTP command. The following information is passed to the user exit:

- The user ID
- The FTP command to be processed

- The command's arguments
- The directory type (MVS or HFS)
- The FILETYPE (SEQ, JES, or SQL)
- The current working directory
- A buffer to hold a modified argument string
- A buffer to hold a 500 reply extension to explain why the exit denied the request
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier
- A 256-byte scratchpad buffer

The user exit enables an installation to modify the arguments of an FTP command or to deny a user from issuing the command. For example, if the server receives a LIST * command, the exit can either deny the command or modify it to LIST 'USER1.*'. If the user exit denies the request by this user to issue this command, one or both of the following replies will be sent to the user. The first reply is optional and is sent only if the user exit returns a string in the 500 reply extension buffer.

```
500-UX-buffercontents
500 User Exit denies Userid userid from using Command command
```

The FTCHKJES user exit (for the FTP server)

FTCHKJES is called if the server is in FILETYPE=JES mode and the client tries to submit a job. The following information is passed to the exit:

- The user ID
- A buffer containing the current JCL statement
- Size of statement in the buffer
- JESLrecl value
- Number of this buffer in current series
- Bytes transferred so far (including this buffer)
- Client identifier (see also session instance identifier)
- JESRecfm value
- FTCHKJES exit-specific workarea (4 bytes)
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier
- A 256-byte scratchpad buffer

The exit can allow or refuse the job to be submitted to the JES internal reader based on any information passed to the exit. For example, the exit can look for a USER= parameter on the JOB statement and check it against the client's user ID. If the remote job submission is denied, the exit sends the user the following reply:

```
550 User Exit refuses this job to be submitted by userid
```

The FTPOSTPR user exit (for the FTP server)

FTPOSTPR is called after execution of the FTP commands RETR, STOR, STOU, APPE, DELE, and RNTO. The following information is passed to the exit:

- The user ID

- Client IP address *
- Client port number *
- The directory type (MVS or HFS)
- The current working directory
- The FILETYPE (SEQ, JES, or SQL)
- Most recent reply code number
- Most recent reply text string
- Current FTP command
- Current CONDDISP setting
- Close reason code
- Name of data set or z/OS UNIX file retrieved or stored
- Bytes transferred
- The socket address structure of the client's control connection
- The socket address structure of the server's control connection
- Session instance identifier
- A 256-byte scratchpad buffer
- The 1-byte description of the confidence-of-successful-completion level assigned to this file transfer

Notes:

1. Fields above marked with an asterisk (*) are valid only for IPv4 addresses, including IPv4-mapped IPv6 addresses.
2. The directory type reflects the current working directory. The MVS data set or z/OS UNIX file that is retrieved or stored can be located in a different type of directory.

The exit allows for post processing at the termination of data transfer functions within the server.

The EZAFCCMD user exit (for the FTP client)

The EZAFCCMD user exit is called by the FTP client before an FTP command is sent to the FTP server. The following information is passed to the exit:

- The number of parameters that are passed to the exit
- The TCP connection ID
- The remote user ID
- The FTP command to be processed
- The arguments of the command
- The directory type (MVS or HFS)
- The z/OS UNIX file type (FILE or FIFO)
- The current local directory
- The fully qualified MVS data set name or z/OS UNIX absolute path name of the client
- The current configuration of the FTP client
- The socket address structure of the control connection for the client
- The socket address structure of the control connection for the server
- The socket address structure of the connection for the SOCKS server
- FTP client application data for the control connection
- A buffer to hold a modified argument string
- A buffer to hold a 71-byte message from the user exit

This exit can inspect an FTP command, modify the arguments of an FTP command, reject an FTP command, or end the FTP client before the command is sent to the server. For example, if the client is sending a LIST * command, the exit can inspect the LIST command, change the command argument to LIST 'USER11.*', reject the LIST command, or end the FTP client.

Restrictions:

- This user exit cannot end the client when the QUIT command is generated by the QUIT subcommand, but the client does end as part of QUIT subcommand processing.
- This user exit cannot reject the QUIT command that is generated by the QUIT subcommand or by the CLOSE subcommand.
- This user exit cannot change the FTP command, but can change the arguments of the FTP command.
- This user exit cannot modify the arguments of the AUTH, EPRT, EPSV, MODE, PBSZ, PORT, PROT, REST, SITE, STRU, TYPE, and XLMT commands.
- This user exit cannot inspect the PASS or ACCT command arguments, but can modify the arguments of these commands.
- This user exit cannot inspect or modify the ADAT command arguments.
- This user exit cannot modify the RETR or STOR command arguments when the transferred MVS data set is a load module.

For more information about the EZAFCCMD user exit, see z/OS Communications Server: IP Configuration Reference.

The EZAFCREP user exit (for the FTP client)

The EZAFCREP user exit is called when the FTP client receives a single-line reply or one line of a multiple-line reply from the server over the control connection. The following information is passed to the user exit:

- The number of parameters that are passed to the user exit
- The TCP connection ID
- The FTP server reply line that is received by the FTP client
- FTP client application data for the current control connection
- A 71-byte buffer to hold a message from the user exit

Use the EZAFCREP user exit to inspect the FTP server reply or to end the FTP client after the FTP client receives a particular line of the reply from the server. If the user exit ends the FTP client, one or both of the following messages are displayed to the user. Message EZA1556I is optional and is displayed only when the user exit returns text in the 71-byte buffer.

```
EZA1556I EZAFCREP message: message text provided by user exit routine
EZA1546I User exit EZAFCREP  module modname ended the FTP client - exit reason code
x'hexadecimal-rsncode' (decimal-rsncode)
```

Restrictions:

- This user exit cannot modify FTP server replies.
- This user exit cannot end the client when the reply is generated by the QUIT subcommand, but the client does end as part of QUIT subcommand processing.
- This user exit cannot end the client when the input to the exit is a reply with a reply code in the range 100–199.

For more information about the EZAFCREP user exit, see z/OS Communications Server: IP Configuration Reference.

Customizing the FTP-to-JES interface for JESINTERFACELevel 2 (optional)

If FTP.DATA does not change the JESINTERFACELEVEL to 2, the FTP server uses the JES interface provided in releases prior to CS for OS/390 V2R10. At this level, the FTP user is allowed to submit jobs to JES, retrieve held output matching their logged-in user ID plus one character, and delete held jobs matching their logged-in user ID plus one character.

If JESINTERFACELevel is set to 2, FTP users have the ability to retrieve and delete any job in the system permitted by the System Authorization Facility (SAF) resource class JESSPOOL. For that reason, JESINTERFACELevel=2 should be specified only if the appropriate JES and SDSF security measures are in place to protect access to JES output. The SAF controls used for JESINTERFACELevel=2 are essentially a subset of those used by SDSF. Therefore, if an installation has customized SAF facilities for SDSF, they are configured for FTP JES level 2.

Note: You are not required to have SDSF to use JESINTERFACELEVEL 2. If you do not use SDSF, you need to create SAF profiles. Both SDSF and JESINTERACELEVEL 2 use the same SAF profile names.

Before customizing the FTP-to-JES interface, complete JES customization. For example, JESJOBS is an SAF class that controls which users can submit jobs to JES. JESSPOOL is the SAF class that controls which users can access output jobs. Customize these SAF classes before beginning customization of the FTP-to-JES interface.

JESSPOOL defines resource names as <nodeid>.<userid>.<jobname>.<Dsid>.<dsname>. An FTP user can delete an output job if they have UPDATE access to the resource that matches their nodeid, userid, and job name. If the FTP user has READ access to the resource, they can list, retrieve, or GET the job output. For more information on JES security, see z/OS JES2 Initialization and Tuning Guide. For more information on the SAPI interface, see z/OS MVS Using the Subsystem Interface.

There are three filters used by the FTP server to control the display of jobs:

- JESSTATUS
- JESOWNER
- JESJOBNAME

SDSF resources are employed for this.

JESSTATUS can be changed by an FTP user with the SITE command to filter jobs in INPUT, ACTIVE, or OUTPUT state. The SDSF resources checked for these states are ISFCMD.DSP.INPUT.jesx, ISFCMD.DSP.ACTIVE.jesx, and ISFCMD.DSP.OUTPUT.jesx, respectively. At login time (USER command), the default value is set to ALL if READ access is allowed to all three classes. Otherwise it attempts to set it to OUTPUT, ACTIVE, and then INPUT if the appropriate READ access is allowed. If no READ access is allowed to any of the classes, JESSTATUS is set to OUTPUT but JESOWNER and JESJOBNAME cannot be changed from the default. In this way, SAF controls can be put in place to limit FTP users to whatever status of jobs an installation requires.

At login time, JESOWNER will have the value of the logged-in user ID. Authority to change JESOWNER is obtained through READ access to RACF profile ISFCMD.FILTER.OWNER. An FTP user who has READ access to ISFCMD.FILTER.OWNER will be allowed to change the JESOWNER parameter with the SITE command.

At login time, JESJOBNAME will have the value of the logged-in user ID plus an asterisk (*). Authority to change JESJOBNAME is obtained through READ access to RACF profile ISFCMD.FILTER.PREFIX. An FTP user who has READ access to ISFCMD.FILTER.PREFIX will be allowed to change the JESJOBNAME parameter with the SITE command.

For example, to allow all users except USER1 to be allowed to change JESOWNER enter the following commands:

```
SETROPTS CLASSACT(SDSF) REFRESH
RDEFINE SDSF (ISFCMD.FILTER.OWNER) UACC(READ)
```

```
PERMIT ISFCMD.FILTER.OWNER ACCESS(NONE) CLASS(SDSF) ID(USER1)
SETROPTS CLASSACT(SDSF) REFRESH
```

For more information on SDSF security, see z/OS SDSF Operation and Customization.

Configuring the FTP server for anonymous FTP (optional)

You can configure the FTP server to allow users to log in anonymously. A user logs in anonymously by logging in as `anonymous` instead of as a user ID defined to the system. To enable users to log in anonymously, code the ANONYMOUS statement in the server FTP.DATA data set.

You can specify three levels of anonymous support on the ANONYMOUSLEVEL statement.

- ANONYMOUSLEVEL 1

  That is, the ANONYMOUS statement is supported. If no operands are specified on the ANONYMOUS statement, the anonymous user needs no password and has access to MVS data sets and the z/OS UNIX file system.
- ANONYMOUSLEVEL 2

  You should not specify ANONYMOUSLEVEL 2; ANONYMOUSLEVEL 2 is provided for migration purposes only. Consider ANONYMOUSLEVEL 3 if ANONYMOUSLEVEL 1 does not meet your anonymous login security requirements.
- ANONYMOUSLEVEL 3

  ANONYMOUSLEVEL 3 is the default.If you specify ANONYMOUSLEVEL 3, the anonymous user cannot issue the USER command to leave anonymous mode, nor can another user issue USER `anonymous` to enter anonymous mode.

  If you specify ANONYMOUSLEVEL 3 and STARTDIRECTORY HFS in FTP.DATA, the following rules apply:

  – The anonymous user's z/OS UNIX file system access is restricted to the anonymous user's home directory and home directory subtrees. The anonymous user's home directory is the home directory of the user ID coded on the ANONYMOUS statement or the home directory of the user ID ANONYMO if you code the ANONYMOUS statement without a user ID.

  – You must create an anonymous directory structure in the z/OS UNIX file system. See “Creating an anonymous directory structure in the z/OS UNIX file system” on page 715 for more information.

  If you specify ANONYMOUSLEVEL 3 and STARTDIRECTORY MVS in FTP.DATA, you must create a shadow copy of the `/usr/sbin/ftpdns` path and file under the home directory of the anonymous user in the z/OS UNIX file system. For more information, see step 2 in “Creating an anonymous directory structure in the z/OS UNIX file system” on page 715

  The ANONYMOUSLEVEL 3 server recognizes additional statements that restrict the anonymous user's access to FTP resources. The following statements are ignored when ANONYMOUSLEVEL is 1 or 2:

  – ANONYMOUSFILEACCESS allows the system programmer to preclude access to either the z/OS UNIX file system or MVS data sets.

  – ANONYMOUSFILETYPEJES, ANONYMOUSFILETYPESQL, and ANONYMOUSFILETYPESEQ control whether the anonymous user can set filetype JES, SQL, or SEQ, respectively.

  – ANONYMOUSHFSFILEMODE defines the mode bits used for files written to the z/OS UNIX file system.

  – ANONYMOUSHFSDIRMODE defines the mode bits used for directories created in the z/OS UNIX file system.

  When ANONYMOUSLEVEL is set to 3, the user's email address is requested in lieu of a password in the following situations:

  – ANONYMOUS is specified without any parameters.

  – ANONYMOUS is specified with user ID/password.

  – ANONYMOUS is specified with user ID/SURROGATE.

You can control the degree of verification of the email address that an anonymous user enters as a password by using the EMAILADDRCHECK keyword in FTP.DATA. See z/OS Communications Server: IP Configuration Reference for details about the EMAILADDRCHECK keyword. The email address entered is logged to the syslog daemon and is also passed to a user exit routine, FTCHKPWD, for user processing.

The FTP server can be defined to process users without passwords by using the ANONYMOUS SURROGATE support. In order to support this, ANONYMOUSLEVEL must be set to 3 in FTP.DATA on the server and BPX.SRV surrogate must be defined in RACF.

z/OS UNIX uses profiles defined to the RACF SURROGAT class to authorize the server to act as a surrogate of a client. Profiles defined to the SURROGAT class are of the form:

```
BPX.SRV.<userid>
```

in which <userid> is the MVS user ID of the user that the server will support without a password.

The following steps are for a sample user ID of the FTP daemon (the user ID associated with the FTP started task procedure) called FTPD with the ability to support user ID GUEST without a password. As you add more servers, you will need to follow similar procedures.

1. Activate the SURROGAT class support in RACF:

   ```
   SETROPTS CLASSACT(SURROGAT)
   ```

   This has to be done only once on the system. The SURROGAT class may already have been set up on your system. If a daemon or server you are running will be using the SURROGAT support heavily, consider using the RACLIST command to keep the SURROGAT profiles in storage. The following example shows how to cache the SURROGAT profiles in storage:

   ```
   SETROPTS RACLIST(SURROGAT)
   ```

2. If the SURROGAT profile is in the RACLIST, any changes to the SURROGAT profiles must be followed by a REFRESH command. To create the SURROGAT class profile for user ID GUEST, issue:

   ```
   RDEFINE SURROGAT BPX.SRV.GUEST UACC(NONE)
   SETROPTS RACLIST(SURROGAT) REFRESH
   ```

   A similar SURROGAT profile is required for each user ID that a server must support without a password.

3. To permit the user ID of the FTP daemon (the user ID associated with the FTP started task procedure), FTPD, to create a security environment for user ID GUEST, issue the PERMIT command:

   ```
   PERMIT BPX.SRV.GUEST CLASS(SURROGAT) ID(FTPD) ACCESS(READ)
   SETROPTS RACLIST(SURROGAT) REFRESH
   ```

Creating an anonymous directory structure in the z/OS UNIX file system

The sample shell script, ftpandir.scp, will create an anonymous directory structure for you, containing required and optional structures. Or, a superuser can create the anonymous directory structure. In this topic, the steps a superuser would follow to create an anonymous directory structure are outlined.

For the following steps, assume that the RACF user ID that is used when an anonymous user logs in is called GUEST, that the HOME directory in that user's OMVS segment in RACF is /u/guest, and that FTP.DATA contains a statement similar to this: ANONYMOUS GUEST

1. Create a bin subdirectory in the anonymous root containing the executable files `ls` and `sh`. This is a required directory. `ls` can be copied from the standard directory. `sh` is part of the standard MVS search order, so you need only create an empty file with the sticky bit.

   The following example shows how to create ls and sh in the user GUEST's home directory:

   ```
   ===> cd /u/guest
   ===> mkdir bin
   ===> chmod 711 bin
   ```

```
===> cd bin

===> cp /bin/ls ls
===> chmod 711 ls
===> touch sh
===> chmod 711 sh
===> chmod +t sh
```

An **ls -al** command should give the following results. Owner and group attributes may be different in your system.

```
# ls -al
total 280
drwx--x--x   2 USER22   0  8192 Sep 21 17:39 .
drwx--x--x   7 USER22   0  8192 Nov  1 14:44 ..
-rwx--x--x   1 USER22   0  126976 Sep 21 17:39 ls
-rwx--x--t   1 USER22   0  0 Sep 21 17:39 sh
```

2. Create a usr/sbin sudirectory of the anonymous root containing the executable file ftpdns. This is a required subdirectory. The file ftpdns can be empty with the sticky bin on.

   The following example is for anonymous user GUEST:

   ```
   ===> cd /u/guest
   ===> mkdir usr
   ===> chmod 711 usr

   ===> cd usr
   ===> mkdir sbin
   ===> chmod 711 sbin
   ===> cd sbin
   ===> touch ftpdns
   ===> chmod 711 ftpdns
   ===> chmod +t ftpdns
   ```

   If you do not configure the subdirectories, bin and usr/sbin, and their contents correctly, the FTP server will not be able to accept anonymous logins and message EZYFT731 will be displayed.

3. Create a dev subdirectory within the anonymous root. This is a required subdirectory. A null file is created in this directory and used during the open of syslog.

   The following example is for anonymous user GUEST:

   ```
   ===> cd /u/guest
   ===> mkdir dev
   ===> chmod 711 usr
   ```

   If you do not have the dev subdirectory, syslog might not open correctly. Messages such as EZA2830I will not be logged out correctly.

4. Set up the public directory structure. This is a required directory.

   This is the directory structure into which you place files that can be downloaded by the anonymous FTP user. It does not have to be named pub; it can be any name you choose. A general convention for anonymous FTP sites is to call it pub:

   ```
    ===> cd /u/guest
   ===> mkdir pub
   ===> cd pub
   ```

   If you want to structure the files you allow to be accessed, you can create multiple subdirectories underneath this directory.

   For simplicity, assume a single level directory, the pub directory. Into this directory you copy the files you want to allow the anonymous user to download:

   ```
   ===> cp /x/y/z/prodinfo1.txt prodinfo1.txt
   ===> cp /x/y/z/prodinfo2.txt prodinfo2.txt
   ===> cd ..
   ```

Make sure that the permission bits are set correctly by using the following shell command when executed in the /u/guest directory. This will set the permission bits of all files in the pub directory and its subdirectories to 755:

```
===> chmod -R 755 pub
```

If your system does not require an incoming or extract directory, the system is configured for anonymous FTP. An **ls -al** command of the pub directory should give the following results:

```
drwxr-xr-x   3 IBMUSER  SYS1  8192 May 13 21:15 .
drwxr-xr-x   6 IBMUSER  SYS1  8192 May 20 14:51 ..
-rwxr-xr-x   1 IBMUSER  SYS1  12 May 11 12:41 prodinfo1.txt
-rwxr-xr-x   1 IBMUSER  SYS1  12 May 11 12:41 prodinfo2.txt
```

5. Set up an incoming directory (optional).

   If you want anonymous users to be able to upload files to your FTP server, you need some additional setup. The objective is to allow an anonymous user to upload a file, but not to allow another anonymous user to download or even be aware of the existence of the file until after an administrative user has verified that the content of the file is acceptable. You do not want your FTP server site to become a store-and-forward site for files of questionable ethical content.

   Positioned at the /u/guest directory, a superuser issues the following shell command:

   ```
   ===> cd /u/guest
   ===> mkdir incoming
   ===> chmod 733 incoming
   ```

   It does not have to be named incoming; it can be any name you choose. A general convention for anonymous FTP sites is to call it incoming.

   The 733 permission bits means that a non-superuser cannot list the content of the incoming directory, but can write a file to it. Because the FTP server enforces a UMASK of 777 when an anonymous user logs in, these files will be written with permission bits 000, which means that they cannot be accessed by the anonymous user or by any other user except a superuser.

   An FTP client user can usually change the UMASK value with a SITE UMASK command or the user can change the permission bits of files they own through a SITE CHMOD command.

   If you define ANONYMOUSLEVEL 3, you can use the ANONYMOUSHFSDIRMODE keyword to set the permission bits of any directory created by an anonymous user, and the ANONYMOUSHFSFILEMODE to set the permission bits of any file created by an anonymous user.

   If you do allow anonymous users to store files on your FTP server, you should ensure that the directory into which these files are stored is a separate z/OS UNIX file system that can fill up without impacting other work on your z/OS system. The best way to do that is to allocate the /u/guest/incoming directory in its own z/OS File System, HFS data set, or Network File System. If an anonymous user uploads large amounts of data to the incoming directory, only this separate z/OS UNIX file system will be filled up. Filling this separate z/OS UNIX file system prevents other anonymous users from storing new files on the server, but will not affect other functions on your system. At a minimum, you should make sure that the incoming directory is not located on the same physical device as your /tmp directory.

6. Set up the extract directory (optional).

   If you need to make files available to certain anonymous users, but not to everyone, you can create a directory that cannot be listed, but files in it can be downloaded if the anonymous user knows the name of the file.

   Positioned at the /u/guest directory, a superuser issues the following shell commands:

   ```
   ===> cd /u/guest
   ===> mkdir extract
   ===> chmod 711 extract
   ```

   It does not have to be named extract; it can be any name you choose. A general convention for anonymous FTP sites is to call it extract.

A superuser can then copy files into this directory, ensure they have permissions of 755, inform the intended anonymous user of the file name, and that user can then log on as anonymous and retrieve the file.

An **ls -al** command at the /u/guest location should give the following result, if you created all four subdirectories:

```
drwxr-xr-x   6 IBMUSER  SYS1 8192 May 20 14:51 .
dr-xr-xr-x   6 IBMUSER  SYS1  0 Jun 10 15:43 ..
drwx--x--x   2 IBMUSER  SYS1  8192 May 11 12:44 bin
drwx--x--x   3 IBMUSER  SYS1  8192 May 11 13:39 extract
drwx-wx-wx   3 IBMUSER  SYS1  8192 May 25 09:35 incoming
drwxr-xr-x   3 IBMUSER  SYS1  8192 May 13 21:15 pub
```

Configure the welcome banner page, login, and directory message (optional)

The FTP server provides support to enable FTP administrators to provide useful information about the site to FTP users. The following FTP.DATA statements are available:

- BANNER
- LOGINMSG
- ANONYMOUSLOGINMSG
- MVSINFO
- ANONYMOUSMVSINFO
- HFSINFO
- ANONYMOUSHFSINFO

You can use the LOGINMSG statement in FTP.DATA to point to a set of messages displayed when a known user logs in to FTP. Similarly, ANONYMOUSLOGINMSG can point to a set of messages displayed when an anonymous user logs in to FTP.

You can use the MVSINFO statement to point to a set of messages displayed when a known user changes the working directory to a particular MVS data set path. Likewise, use the ANONYMOUSMVSINFO statement to point to a set of messages displayed when an anonymous user changes working directory to a particular MVS data set path.

You can use the HFSINFO statement to point to a set of messages displayed when a client changes the working directory to a particular z/OS UNIX directory. Likewise, use the ANONYMOUSHFSINFO statement to point to a set of messages displayed when an anonymous user changes working directory to a particular z/OS UNIX directory.

Using magic cookies to represent information

The content of all the informational messages may include a predefined set of magic cookies, which are substituted by the FTP server before the data is sent to the FTP client. The following magic cookies are supported:

- %T — Local time
- %C — Current working directory
- %E — The email address of the FTP server administrator
- %R — Remote host name
- %L — Local host name
- %U — Username (logged in user)

If %R is used, a long delay in login processing might occur as the FTP server will issue a DNS query to resolve the remote host IP address. In order to use %E, the ADMINEMAILADDR keyword must be specified in the server FTP.DATA configuration file.

Configuring the FTP server to log session (user ID) activity

You can configure the FTP server to write log messages for trace activity related to individual sessions by coding the FTPLOGGING or ANONYMOUSFTPLOGGING statements in FTP.DATA. If you have configured syslogd for FTP, the log messages appear in syslog.

Table 34 on page 719 shows the FTP log message written for various activities. You can identify an FTP log message by the ID=*sessionID* parameter in the message. Each login session is assigned a session ID that can be used to identify all log messages related to that session. For more information about EZYF*xxxx* messages, see z/OS Communications Server: IP Messages Volume 3 (EZY).

Table 34. EZYFxxxx messages

FTP log message	**Activity**
EZYFS50I	A client connected to the FTP daemon.
EZYFS51I	A client connection to the FTP daemon failed.
EZYFS52I	An FTP session ended.
EZYFS54I	The server accepted the security mechanism. The connection with the client is now protected.
EZYFS55I	The server rejected the security mechanism. The connection with the client is not protected.
EZYFS56I	A client logged into the server.
EZYFS57I	A client login to the server failed.
EZYFS58I	The server denied the client access to an MVS data set.
EZYFS59I	The server denied the client access to a z/OS UNIX file.
EZYFS60I, EZYFS61I, EZYFS62I	The server successfully allocated an MVS data set.
EZYFS63I, EZYFS64I, EZYFS65I	The server could not allocate an MVS data set.
EZYFS67I	The server successfully allocated an MVS data set.
EZYFS68I, EZYFS69I	The server could not allocate a z/OS UNIX file.
EZYFS70I, EZYFS74I, EZYFS75I	The server deallocated an MVS data set.
EZYFS71I, EZYFS72I, EZYFS73I	The server detected an error while deallocating an MVS data set.
EZYFS77I	The server deallocated a z/OS UNIX file.
EZYFS78I, EZYFS79I	The server detected an error while deallocating a z/OS UNIX file.
EZYFS80I	The server sent a reply to the client after data transfer.
EZYFS81I	The server finished processing an MVS data set transfer.
EZYFS82I	The server finished processing a z/OS UNIX data set transfer.
EZYFS83I	The server stored a data set or file into its file system.
EZYFS84I	The server sent a data set or file to the client (or to a server in the case of a proxy transfer).
EZYFS85I	The server finished returning job output to the client.
EZYFS86I	The server assigned a confidence of success level to the completed data transfer.
EZYFS91I	The server submitted a job for the client.

Table 34. EZYFxxxx messages (continued)	
FTP log message	**Activity**
EZYFS92I	The server returned a SQL report.
EZYFS95I	An abend occurred while the server was transferring data.

Configuring to send detailed login failure replies to an FTP client (optional)

The FTP server returns minimal information to the client when the PASS command fails. However, you can configure the FTP server to send additional information by coding ACCESSERRORMSGS TRUE in FTP.DATA. This directs the server to reply to the client with detailed login failure data. The reply might report server errors, such as failing function calls with diagnostic return codes. It might report user errors, such as an expired or incorrect password or password phrase, an unknown user ID, or a revoked user ID. You should not code ACCESSERRORMSGS TRUE in FTP.DATA if you do not want to share this type of information with users logging in to FTP.

You can capture the same information in syslog by coding FTPLOGGING TRUE and ANONYMOUSFTPLOGGING TRUE in FTP.DATA. You can also turn on the DEBUG option called ACC to log the error messages in the syslog. For more information on coding FTP.DATA statements, see z/OS Communications Server: IP Configuration Reference.

Install the SQL query function (optional) and access the Db2 modules

To use FTP to perform SQL queries, bind a plan that allows FTP to invoke the package EZAFTPMQ in collection EZAFTPMQ, and grant execution privileges for that plan to PUBLIC. You can specify the name of the plan using the DB2PLAN keyword in FTP.DATA, or the default is EZAFTPMQ. This FTP facility performs only SELECT operations on the Db2 tables; it does not perform UPDATE, INSERT, or DELETE operations.

Requirement: If secondary authorization for SQL queries is required, you must modify the DSN3SATH sample exit shipped by Db2. The exit returns the primary AUTHID for requests that originate from the FTP server.

The following sample job is provided in the FTOEBIND member of the SEZAINST data set; you can use it to enable the FTP server and client to perform SQL queries.

```
//FTPSETUP JOB FTPSETUP,
//            CLASS=A,
//            NOTIFY=&SYSUID
//*************************************************************
//*
//*   File name:                 tcpip.SEZAINST(FTOEBIND)
//*   SMP/E distribution name:   EZAFTPAB
//*
//*    Licensed Materials - Property of IBM
//*    This product contains "Restricted Materials of IBM"
//*    5647-A01     Copyright IBM Corp.  1997, 2011
//*    All rights reserved.
//*    US Government Users Restricted Rights -
//*    Use, duplication or disclosure restricted by GSA ADP Schedule
//*    Contract with IBM Corp.
//*    See IBM Copyright Instructions.
//*
//*  This JCL binds a plan that allows the application to ability the
//*  invoke the package EZAFTPMQ in collection EZAFTPMQ.
//*  The JCL allows execution of the plan EZAFTPMQ by PUBLIC.
//*
//*  The FTP server and client use this plan. (See
//*  Usage note #7)
//*
//*
//*  Usage notes:
```

```
//*
//*    1.  You must execute this job from a user ID that has
//*        the authority to bind the EZAFTPMQ plan.
//*
//*    2.  Change the STEPLIB DD statement in the FTPBIND and
//*        FTPGRANT steps to reflect the DB2 DSNLOAD data set.
//*
//*    3.  Change the DB2 sybsystem name in the FTPBIND and
//*        FTPGRANT steps from SYSTEM(xxx) to the
//*        installation defined DB2 subsystem name.
//*
//*    4.  Change the library parameter in the FTPBIND step from
//*        TCPIP.SEZADBRM to the installation defined TCPIP
//*        SEZADBRM library.
//*
//*    5.  Change the plan name in the FTPGRANT step from
//*        DSNTIAYY to reflect the plan associated with the
//*        program DSNTIAD.
//*
//*    6.  Change the library parameter in the FTPGRANT step
//*        from xxxxxx.RUNLIB.LOAD to reflect the library
//*        where the DSNTIAD program resides.
//*
//*    7.  You can bind the DBRM to a package and a plan name
//*        other than EZAFTPMQ by changing the plan specified
//*        in the FTPBIND and FTPGRANT steps.
//*
//*************************************************************
//FTPBIND  EXEC PGM=IKJEFT01,DYNAMNBR=20
//STEPLIB  DD   DSN=xxxxxx.DSNLOAD,DISP=SHR
//SYSTSPRT DD   SYSOUT=*
//SYSPRINT DD   SYSOUT=*
//SYSOUT   DD   SYSOUT=*
//SYSTSIN  DD   *
 DSN  SYSTEM(xxx)
 BIND CURRENTDATA(NO) -
      EXPLAIN(NO) -
      ISOLATION(CS) -
      LIBRARY('TCPIP.SEZADBRM') -
      MEMBER(EZAFTPMQ) -
      PACKAGE(EZAFTPMQ) -
      RELEASE(COMMIT) -
      VALIDATE(RUN)
 BIND ACQUIRE(USE) -
      ACTION(REPLACE) -
      CACHESIZE(1024) -
      CURRENTDATA(NO) -
      EXPLAIN(NO) -
      ISOLATION(CS) -
      NODEFER(PREPARE) -
      PKLIST(EZAFTPMQ.EZAFTPMQ) -
      PLAN(EZAFTPMQ) -
      RELEASE(COMMIT) -
      VALIDATE(RUN) -
      RETAIN
 END
//*
//FTPGRANT EXEC PGM=IKJEFT01,DYNAMNBR=20
//STEPLIB  DD   DSN=xxxxxx.DSNLOAD,DISP=SHR
//SYSTSPRT DD   SYSOUT=*
//SYSPRINT DD   SYSOUT=*
//SYSOUT   DD   SYSOUT=*
//SYSTSIN  DD   *
 DSN  SYSTEM(xxx)
 RUN  PROGRAM(DSNTIAD) -
      PLAN(DSNTIAYY) -
      LIBRARY('xxxxxx.RUNLIB.LOAD')
 END
//SYSIN    DD   *
 GRANT EXECUTE ON PLAN EZAFTPMQ TO PUBLIC;
//*
```

Accessing Db2 modules

The FTP server or client loads 3 Db2 modules into storage to perform an SQL query. These modules are:

- DSNALI
- DSNHLI2

- DSNTIAR

The modules are usually found in the Db2 load library with the suffix DSNLOAD. The Db2 administrator or system programmer should add the DSNLOAD library to the LINKLIST to ensure FTP has access to this library.

Another way to ensure access is to add the DSNLOAD library to the FTP STEPLIB. For the FTP server this means the JCL used to start the FTP server has a STEPLIB DD statement referring to the DSNLOAD library or, if the FTP daemon is started from the z/OS shell, the STEPLIB environment variable is set. For the FTP client, this means a TSO CLIST must allocate the DSNLOAD library as the STEPLIB.

If the FTP client is to be run from a batch job to perform SQL queries, the DSNLOAD library must be added to the STEPLIB DD statement for the batch job.

Usage notes:

To allow FTP access to multiple levels of Db2, link to the libraries that contain the lowest level of Db2 to be accessed.

FTP.DATA updates for SQL query function

To obtain FTP.DATA updates for the SQL query function, follow these steps:

1. Set the FTP.DATA DB2 statement to specify the name of the Db2 subsystem.
2. Set DB2PLAN to specify the Db2 plan to be used by the FTP server.
3. Set the SPREAD statement to specify whether SQL output is in spreadsheet format.
4. Set SQLCOL to specify the column headings of the output data.

Verifying the FTP server

If FTP is in the autolog list and the TCP/IP address space is restarted, FTP should start automatically. For other cases, it should be started manually. To do this, go to the MVS console and enter the following command:

```
S FTPD
```

Note: This command assumes the FTP procedure name is FTPD.

If the FTP server startup is complete, the following message should be seen on the MVS console:

```
EZY2702I Server-FTP: Initialization completed at 17:37:29 on  12/17/99.
```

If the message is not seen, a message explaining why FTP did not start up will appear in SYSLOG. Even if the above message is issued, it would be beneficial to inspect SYSLOG for warning messages issued during FTP initialization. EZY2700I displays the port FTP uses as the control port, the port it listens to for incoming connections from clients. In this example, FTP is listening to standard port 21.

The file syslog uses is defined in /etc/syslog.conf. The statement `daemon.info /tmp/daemon.log` directs SYSLOGD to save all the daemon messages in /tmp/daemon.log. Below is an example of output error messages.

```
EZYFT18I Using catalog '/usr/lib/nls/msg/C/ftpdmsg.cat' for FTP messages.
EZY2697I IBM FTP CS V1R10 21:04:56 on 04/17/08
EZY2640I Using dd:SYSFTPD=USER1.FTP.DATA for local site configuration parameters
EZYFT46E Error in dd:SYSFTPD file: line 4 near column 9.
EZY2636E SMFLOGN value not specified.
EZYFT46E Error in dd:SYSFTPD file: line 5 near column 8.
EZY2636E SMFREN value not specified.
EZYFT47I dd:SYSFTPD file, line 21: Ignoring keyword "EXTENSIONS REST_STREAM".
EZYFT47I dd:SYSFTPD file, line 29: Ignoring keyword "CTRLCONN".
EZYFT21I Using catalog '/usr/lib/nls/msg/C/ftpdrply.cat' for FTP replies.
EZYFT26I Using 7-bit conversion derived from 'ISO8859-1' and 'IBM-1047' for the control connection.
EZYFT33I Unable to open DDNAME 'SYSFTSX' for the data connection: EDC5129I No such file or directory.
EZYFT31I Using //'TPOUSER.STANDARD.TCPXLBIN' for FTP translation tables for the data connection.
EZYFT09I system information for VIC135: z/OS version 1 release 10 (2094)
EZY2700I Using port FTP control (21)
EZY2701I Inactivity time is 0
```

```
EZY2702I Server-FTP: Initialization completed at 21:05:57 on 04/17/08.
EZYFT41I Server-FTP: process id 16777255, server job name FTPD11
```

Verifying the FTP client

To verify that the FTP client works correctly, log onto TSO and issue the NETSTAT HOME command, or issue NETSTAT –h from the z/OS UNIX shell. These commands will show the interface addresses that are known to the system. Below is an example of the output from NETSTAT HOME:

```
MVS TCP/IP NETSTAT CS V1R9       TCPIP Name: TCPCS           14:22:17
Home address list:
LinkName:   OSAQDIO6L
  Address:  9.67.115.13
    Flags:  Primary
LinkName:   LSAMEH
  Address:  9.1.1.1
    Flags:
LinkName:   LOOPBACK
  Address:  127.0.0.1
    Flags:
Address:  fe80::9:6b00:671a:586
    Type:   Link_Local
    Flags:  Autoconfigured
IntfName:   V6SAMEH
  Address:  1::8
    Type:   Global
    Flags:
IntfName:   V6VIRT
Address:  2::55
    Type:   Global
    Flags:
IntfName:   LOOPBACK6
  Address:  3::1
    Type:   Global
    Flags:
  Address:  ::1
    Type:   Loopback
    Flags:
```

To invoke the FTP client, use any address shown on the NETSTAT HOME address list. The first example below shows how you could log in to the FTP server at 9.67.115.13 using a batch job (the output of the batch job is not shown). The second example shows logging in to the FTP server at 9.67.113.37 from the TSO environment.

```
//FTPBATCH JOB FTPUSER,
// USER=USER1,PASSWORD=TCPSUP
//BATCH  EXEC PGM=FTP
//OUTPUT  DD SYSOUT=*
//INPUT DD *
  9.67.115.13
  USER10 tcpusr
  SITE FILE=SEQ
  QUIT
//*
```

```
Using 'USER1.FTP.DATA' for local site configuration parameters.
IBM FTP CS V1R9
FTP: using TCPCS
Connecting to: vic135.tcp.raleigh.ibm.com 9.67.113.37 port: 21.
220-FTPD1 IBM FTP CS V1R9 at vic135, 19:07:34 on 2006-10-08.
220 Connection will close if idle for more than 5 minutes.
>>> FEAT
211- Extensions supported
 AUTH TLS
 PBSZ
 PROT
211 End
NAME (vic135:USER1):

user1
NAME (vic135:USER1):
>>> USER USER1
331 Send password please.
```

```
PASSWORD:

>>> PASS
230 USER1 is logged on.  Working directory is "/".
Command:
```

Verifying FTP.DATA statements

You can verify many FTP.DATA statements by using the FTP client STAtus and LOCSTat subcommands. The output from the STAtus and LOCSTat subcommands depends on the client and server copy of FTP.DATA at each installation. The following sample shows the output of one system:

```
stat
 EZA1701I >>> STAT
  211-Server FTP talking to host ::ffff:10.1.2.3, port 1026
  211-User: USER1  Working directory: USER1.
  211-The control connection has transferred 115 bytes
  211-There is no current data connection.
  211-The next data connection will be actively opened
  211-to host ::ffff:10.1.2.3, port 1026,
  211-using Mode Stream, Structure File, type ASCII, byte-size 8
  211-Automatic recall of migrated data sets.
  211-Automatic mount of direct access volumes.
  211-Auto tape mount is allowed.
  211-Inactivity timer is set to 300
  211-Timer FTPKEEPALIVE is set to 0
  211-Timer DATAKEEPALIVE is set to 0
  211-Timer DSWAITTIME is set to 0
  211-Timer FIFOOPENTIME is set to 60
  211-Timer FIFOIOTIME is set to 20
  211-VCOUNT is 59
  211-ASA control characters in ASA files opened for text processing will be
  211- transferred as ASA control characters.
  211-Trailing blanks are removed from a fixed format data set when it is
  211- retrieved.
  211-Data set mode.  (Do not treat each qualifier as a directory.)
  211-ISPFSTATS is set to FALSE
  211-Primary allocation 1 track.  Secondary allocation 1 track.
  211-Partitioned data sets will be created with 27 directory blocks.
  211-FileType SEQ (Sequential - default).
  211-Number of access method buffers is 5
  211-RDWs from variable format data sets are discarded.
  211-Records on input tape are unspecified format
  211-SITE DB2 subsystem name is DB2
  211-Data not wrapped into next record.
  211-Tape write is not allowed to use BSAM I/O
  211-Truncated records will not be treated as an error
  211-JESLRECL is 80
  211-JESRECFM is Fixed
  211-JESINTERFACELEVEL is 1
  211-Server site variable JESTRAILINGBLANKS is set to TRUE
  211-Confidence level in data transfers is neither checked nor reported
  211-ENcoding is set to SBCS
  211-MBdataconn codeset names: UTF-16,UTF-16
  211-Outbound SBCS ASCII data uses CRLF line terminator
  211-Outbound MBCS ASCII data uses CRLF line terminator
  211-Server site variable MBREQUIRELASTEOL is set to TRUE
  211-Server site variable UNICODEFILESYSTEMBOM is set to ASIS
  211-Server site variable UNIXFILETYPE is set to FILE
  211-DBSUB is set to FALSE
  211-Server site variable EXTDBSCHINESE is set to TRUE
  211-SBSUB is set to FALSE
  211-SBSUBCHAR is set to SPACE
  211-SMS is active.
  211-New data sets will be catalogued if a store operation ends abnormally
  211-Single quotes will override the current working directory.
  211-UMASK value is 027
  211-Process id is 19
  211-Checkpoint interval is 0
  211-Server site variable RESTPUT is set to TRUE
  211-Server site variable REMOVEINBEOF is set to FALSE
  211-Authentication type: None
  211-TLS security is supported at the DRAFT level
  211-Server site variable READVB is set to LE
  211-Port of Entry resource class for IPv4 clients is: TERMINAL
  211-Record format VB, Lrecl: 256, Blocksize: 6233
  211-Server site variable EATTR is set to OPT
  211-Server site variable DSNTYPE is set to BASIC
```

```
  211-Server site variable LISTSUBDIR is set to TRUE
  211 *** end of status ***


locstat
  Trace: FALSE, Send Port: TRUE
  Send Site with Put command: TRUE
  Connected to:10.1.2.3, Port: FTP control (21), logged in
  Local Port: 1026
  Data type:a, Transfer mode:s, Structure:f
  Automatic recall of migrated data sets.
  Automatic mount of direct access volumes.
  Data set mode. (Do not treat each qualifier as a directory.)
  ISPFSTATS is set to FALSE
  Primary allocation 1 track, Secondary allocation 1 track.
  Partitioned data sets will be created with 27 directory blocks
  FileType is SEQ (Sequential - the default).
  Number of access method buffers is 5.
  Confidence level of data transfers is neither checked nor reported
  ENcoding is set to SBCS
  MBdataconn codeset names: UTF-16,UTF-16
  Outbound SBCS ASCII data uses CRLF line terminator
  Outbound MBCS ASCII data uses CRLF line terminator
  local site variable UNICODEFILESYSTEMBOM is set to ASIS
  local site variable MBREQUIRELASTEOL is set to TRUE
  local site variable REMOVEINBEOF is set to FALSE
  DBSUB is set to FALSE
  SBSUB is set to FALSE
  SBSUBCHAR is set to SPACE
  RDW's from VB/VBS files are discarded.
  Records on input tape are unspecified format
  DB2 subsystem name is DB2
  Volid of Migrated Data Sets is MIGRAT
  Trailing blanks in records read from RECFM F datasets are discarded.
  Record format: VB, Lrecl: 256, Blocksize: 6233.
  Data not wrapped into next record.
  Truncated records will not be treated as an error.
  Checkpoint interval is 0
  Checkpoint data set will be opened for GET
  CHKPTPrefix uses Home to determine the HLQ of the FTP.CHECKPOINT file.
  No automatic mount of tape volumes.
  CCONNTIME is 30
  DATACTTIME is 120
  DCONNTIME is 120
  INACTTIME is 120
  MYOPENTIME is 60
  FTPKEEPALIVE is 0
  local site variable DATAKEEPALIVE is set to 0
  local site variable DSWAITTIME is set to 0
  VCOUNT is 59
  Prompting: ON, Globbing: ON
  ASA control characters transferred as ASA control characters
  New data sets catalogued if a store operation terminates abnormally
  Single quotes will override the current working directory
  UMASK value is 027
  Data connections for the client are not firewall friendly.
  local site variable EPSV4 is set to TRUE
  local site variable SECUREIMPLICITZOS is set to TRUE
  local site variable PASSIVEIGNOREADDR is set to FALSE
  local site variable TLSRFCLEVEL is set to RFC4217
  local site variable READVB is set to LE
  local site variable EXTDBSCHINESE is set to TRUE
  local site variable LISTSUBdir is set to TRUE
  local site variable PROGRESS is set to 10
  local site variable SEQNUMSUPPORT is set to FALSE
  local site variable UNIXFILETYPE is set to FILE
  local site variable FIFOIOTIME is set to 20
  local site variable FIFOOPENTIME is set to 60
  local site variable EATTR is set to SYSTEM
  local site variable DSNTYPE is set to BASIC
  Authentication mechanism: None
  Tape write is not allowed to use BSAM I/O
  Using /etc/ftp.data for local site configuration parameters.
```

Verifying anonymous, banner, and other optional configuration information

Depending on your installation's choices for anonymous level, banner support chosen, exits, and so on, verification of support output will differ. To verify anonymous configuration at a particular installation, log in as anonymous and verify the behavior is as expected. For example, if EMAILADDRCHECK FAIL is specified in FTP.DATA, try to log in as anonymous using an incorrect email address as password. To verify banner support, login and verify the banners are displayed as expected. Below is a sample of FTP.DATA and FTP client output for one such installation.

```
; BANNER STUFF
EMAILADDRCHECK FAIL
BANNER USER1.TEST1
ADMINEMAILADDR FTPADMIN@MYSYSTEM.COM
; ANONYMOUS STUFF
ANONYMOUSLEVEL 3
STARTDIRECTORY HFS
```

```
ftp 9.67.113.63
 IBM FTP CS V1R9  2006 349 01:35 UTC
 FTP: using TCPCS
 Connecting to:  9.67.113.63 port: 21.
 220-FTPD1 IBM FTP CS V1R9 at HOSTA, 19:07:34 on 2006-01-08.
 220-You have just read 'USER1.TEST1'
 220-ADMINEMAILADDRESS is FTPADMIN@MYSYSTEM.COM
 220 Connection will not timeout.
 NAME (9.67.113.63:USER4):
anonymous no-email-pw
 >>> USER anonymous
 331 Send password please.
 >>> PASS
 530 PASS command failed.
 Command:
```

Verifying the FTP-JES interface (optional)

As with the other optional configuration information, FTP-JES support can best be verified by logging in and confirming the FTP.DATA parameters chosen. To verify JES support, a simple batch job can be created if the JESINTERFACELEVEL is set to the security requirements of an installation. Below is the batch job and FTP client output for JESINTERFACELEVEL 2.

```
EDIT       USER1.FTP.JCL.TEST                                Columns 00001 00072
Command ===>                                                    Scroll ===> CSR
****** ***************************** Top of Data ******************************
000100 //JOBTEST  JOB MSGCLASS=H,MSGLEVEL=(1,1),CLASS=A,
000200 //         USER=USER1
000300 //STEP1    EXEC PGM=IEBGENER
000400 //OBJTMP1  DD DSN=&PRLOBJ,DISP=(NEW,PASS,DELETE),
000500 //            SPACE=(CYL,(1,1,10)),
000600 //            DCB=(RECFM=FB,LRECL=80,BLKSIZE=800)
000700 //SYSPRINT DD SYSOUT=A
000800 //SYSUT1   DD DSN=SYS1.PROCLIB(JES2),DISP=SHR
000900 //SYSIN    DD DUMMY
001000 //SYSUT2   DD SYSOUT=H
001100 //
001200 //         EXEC PGM=IEFBR14
****** **************************** Bottom of Data ****************************

site file=jes jesjobname=jobtest jesowner=* jesstatus=all
 EZA1701I >>> SITE file=jes jesjobname=jobtest jesowner=* jesstatus=all
 200 Site command was accepted
 EZA1460I Command:
put 'user1.ftp.jcl.test'
 EZA1701I >>> SITE FIXrecfm 80 LRECL=80 RECFM=FB BLKSIZE=32720
 200 Site command was accepted
 EZA1701I >>> PORT 127,0,0,1,4,12
 200 Port request OK.
 EZA1701I >>> STOR 'user1.ftp.jcl.test'
 125 Sending Job to JES internal reader FIXrecfm 80
 250-It is known to JES as JOB00076
 250 Transfer completed successfully.
```

```
 EZA1617I 984 bytes transferred in 0.005 seconds.  Transfer rate 196.80 Kbytes/
sec.
 EZA1460I Command:
dir j76
 EZA1701I >>> PORT 127,0,0,1,4,13
 200 Port request OK.
 EZA1701I >>> LIST j76
 125 List started OK for JESJOBNAME=JOBTEST, JESSTATUS=ALL and JESOWNER=*
 EZA2284I JOBNAME  JOBID    OWNER    STATUS CLASS
 EZA2284I JOBTEST  JOB00076 USER1    OUTPUT A        RC=0000
 EZA2284I          ID  STEPNAME PROCSTEP C DDNAME   BYTE-COUNT
 EZA2284I          001 JESE              H JESMSGLG      1084
 EZA2284I          002 JESE              H JESJCL        1023
 EZA2284I          003 JESE              H JESYSMSG      1143
 EZA2284I          004 STEP1             H SYSUT2         741
 EZA2284I          005 STEP1             A SYSPRINT       209
 EZA2284I 5 spool files
 250 List completed successfully.
 EZA1460I Command:
```

Chapter 13. The resolver

The resolver acts on behalf of programs as a client to perform the following functions:

- Access name servers to provide name-to-address or address-to-name resolution
- Allocate and read the TCPIP.DATA file
- Establish TCP/IP stack affinity for certain socket APIs
- Provide protocol and services information

To resolve the query for the requesting program, the resolver uses information that it obtains from the following sources:

- Available name servers
- The DNS response information that has been cached locally (when system-wide caching is enabled)
- Local definitions, such as /etc/hosts, /etc/ipnodes, HOSTS.SITEINFO, HOSTS.ADDRINFO, and ETC.IPNODES

The TCPIP.DATA statements control how (and if) the resolver uses name servers. For detailed information about TCPIP.DATA configuration statements, see z/OS Communications Server: IP Configuration Reference.

Requirement: The resolver address space must be started before any application or TCP/IP stack resolver calls can occur.

DNS overview

While TCP/IP applications refer to host computers by their IP addresses, it is easier to use host names. To enable the use of host names in a network, the Domain Name System (DNS) translates host names to IP addresses. Mapping must be consistent across the network to ensure interoperability. DNS provides the host name-to-IP address mapping through network server hosts called *domain name servers*. For detailed information about name servers, see "Domain name servers" on page 730. DNS can also provide other information about server hosts and networks such as the TCP/IP services available at a server host and the location of domain name servers in a network.

DNS organizes the hosts in a network into domains. A *domain* is a group of hosts that share the same name space in the domain hierarchy and are usually controlled within the same organization. Domains are arranged in a hierarchy. A special domain known as the *root domain* exists at the top of the hierarchy. The root domain servers store information about server hosts in the root domain and the name servers in the delegated, *top-level* domains, such as `com` (commercial), `edu` (education), and `mil` (military). The name servers in the top-level domain, in turn, store the names of name servers for their delegated domains, and so on.

The complete name of a host, also known as the *fully qualified domain name* (FQDN), is a series of labels separated by dots or periods. Each label represents an increasingly higher domain level within a network. The complete name of a host connected to one of the larger networks generally has more than one subdomain, as shown in the following examples:

```
host1.subdomain2a.subdomain2.rootdomain
user4720.eng.mit.edu
```

A domain name server requires the FQDN. The client resolver combines the host name with the domain name to create the FQDN before sending the name resolution request to the domain name server.

DNS also provides IP address-to-host name mapping. The DNS defines a special domain called `in-addr.arpa` to translate IPv4 addresses to host names, and the `ip6.int` and `ip6.arpa` domains for IPv6 address-to-host name translation. This kind of mapping is useful for producing output (host names) that is easy to read. An in-addr.arpa name is composed of the reverse octet order of an IP address concatenated with the `in-addr.arpa` string. For example, a host named `Host1` has

© Copyright IBM Corp. 2000, 2022

9.67.43.100 as an IP address. The in-addr.arpa domain translates the Host1 IP address 9.67.43.100 to 100.43.67.9.in-addr.arpa.

A system administrator can name the host systems and domains in the local, private network with any name you want, but to link with name servers in a public network like the Internet, you need to determine which domain you want to be in (which parent domain) and then contact the registrar in that domain to register the names and IP addresses of your name servers. This ensures that queries from outside the domain being defined can be answered by this name server if need be.

Note: Contact the InterNetwork Information Center (InterNIC) for more information about Internet registration. You can contact InterNIC by pointing your web browser at http://www.internic.net.

Domain names

The DNS uses a hierarchical naming convention for naming hosts. Each host name is composed of domain labels separated by periods. Local network administrators have the authority to name local domains within an intranet. Each label represents an increasingly higher domain level within an intranet. The fully qualified domain name of a host connected to one of the larger intranets generally has one or more subdomains:

- *host.subdomain.subdomain.rootdomain*
- *host.subdomain.rootdomain*

Domain names often reflect the hierarchy level used by network administrators to assign domain names. For example, the domain name eng.mit.edu. is the fully qualified domain name, where eng is the host, mit is the subdomain, and edu is the highest level domain (root domain).

Figure 100 on page 730 is an example of the DNS used in the hierarchy naming structure across an intranet.

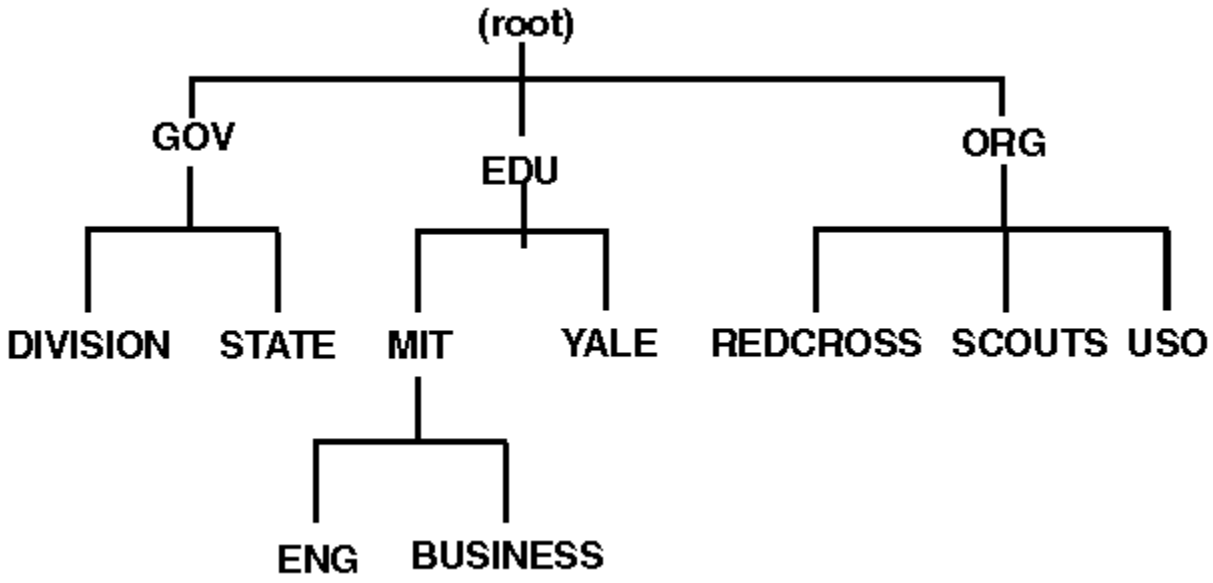


Figure 100. Hierarchical naming tree

You can refer to hosts in your domain by host name only; however, a name server requires a fully qualified domain name. The local resolver appends the domain name before sending the query to the Domain Name Server for address resolution.

Domain name servers

Domain name servers are designated network nodes that maintain a database of information about all nodes in some part of the domain name space, called a *zone*. A name server is said to be *authoritative* for

its zone. A zone consists of the resources within a single domain (for example, commercial or `.com`) or subdomain (for example, `raleigh.ibm.com`). Typically, a zone is administered by a single organization or individual. The complete database is not kept by any one name server on a network. A name server is authoritative only within its zone of authority.

All host systems in a given zone share the same higher level domain name (for example, `host1.raleigh.ibm.com`, `host2.raleigh.ibm.com`, `host3.raleigh.ibm.com`, and so on). As system administrator, you create a zone of authority by listing all the host systems in your zone in the database file of the name server that is authoritative for the zone.

If a domain name server receives a query about a host for which it has information in its database or in its cache, it performs the name resolution and returns all the address records associated with the host to the client. Some hosts (for example, routers or gateways between two or more networks) might have more than one IP address.

Alternatively, the name server can query other name servers for information. This process is called *iterative resolution*. The local name server successively queries other name servers, each of which responds by referring the local name server to a remote name server that is closer to the name server authoritative for the target domain. Finally, the local name server queries the authoritative name server and gets an answer. If the information about a requested host name does not exist or if a name server does not know where to go for the information, it sends a negative response back to the client.

There are multiple name server modes in the DNS:

- Authoritative
  - Master (primary)
  - Secondary
- Caching-only servers
- Forwarders
- Stealth

A single server can perform multiple functions. For example, it can be a primary server and a secondary server for different zones. The purpose of having these different kinds of servers is to provide redundancy (in case of system failure), to distribute the workload among multiple servers, to speed up the name-resolution process, and to provide flexibility in network design. In addition to being an authoritative or caching-only server, a name server can be defined to contact only a specific set of name servers if queries cannot be resolved locally (through the use of forwarders).

The following subtopics discuss authoritative servers, caching-only servers, and forwarding.

Authoritative servers

An authoritative server is the authority for its zone. It queries and is queried by other name servers in the DNS. The data it receives in response from other name servers is cached. Authoritative servers are not authoritative for cached data.

There are two types of authoritative servers: master (primary) and secondary. Each zone must have only one master name server, and it should have at least one secondary name server for backup purposes to minimize dependency on a particular node. Calling a particular name server a master or secondary server is misleading. Any given name server can take on either or both roles, as defined by the conf file.

The zone data updates and maintenance are reflected in the master name server and the changes are then reflected in secondary name servers. Both master and secondary name servers are authoritative for a zone.

The zones of authority are arranged in a hierarchy based on the domain origin components. A special zone known as the *root* exists at the top of the domain name hierarchy in a network. The root zone contains a list of all the root servers. For example (see Figure 100 on page 730), in the Internet, the root name servers store information about nodes in the root domain, and information about the delegated domains, such as `com` (commercial), `edu` (education), and `mil` (military). The root name servers store the names

of name servers for each of these domains, which in turn store the names of name servers for their delegated subdomains.

TCP/IP applications contact a name server whenever it is necessary to translate a domain name into an IP address, or when information is required about a domain. The name server performs the translation if it has the necessary information. If it does not have the necessary information, the name server can contact other name servers, which in turn can contact other name servers. This process is called a *recursive query*. Alternatively, a name server can return the address of another name server that might hold the requested information. This is called a *referral response* to a query. Name server implementations must support referrals, but are not required to perform recursive queries. See “Resolvers” on page 733 for more information about query responses.

Master name servers

A master name server maintains all the data for its zone. Static resources are kept in database files called *domain data files*. Master name servers can also receive zone updates dynamically.

Secondary name servers

A secondary name server acts as an alternate to the master name server if the master name server becomes unavailable or overloaded. The secondary name server receives zone data directly from the master name server in a process called *zone transfer*. Zone transfers occur only when data changes. Some name servers support zone change notification, which results in the zone transfer happening quickly after the data changes. Other name servers poll for changed data based on the refresh interval in the Start of Authority (SOA) resource record. For a description of the SOA resource record, see z/OS Communications Server: IP Configuration Reference. A secondary name server, like a master name server, is authoritative for a zone.

Caching-only servers

All name servers cache (store) the data they receive in response to a query. A caching-only server, however, is not authoritative for any domain. Responses derived from cached information are flagged in the response. When a caching-only server receives a query, it checks its cache for the requested information. If it does not have the information, it queries a local name server or a root name server, passes the information to the client, and caches the answer for future queries. The names and addresses of the root name servers are acquired from the servers listed in the hints file, the name and file path of which are specified in the name server's configuration file.

You can manually configure a name server to create a large cache of responses to queries that are frequently requested and reduce the number of queries made to master servers. The name server that you configure as a caching-only server stores data for a period of time determined by the time-to-live (ttl) value, and the cached information is lost if the name server is restarted.

Tip: As an alternative to manually configuring a caching-only server, you can use the cache that the resolver creates. The resolver cache is enabled by default and typically provides better system performance than using a caching-only name server that you have configured manually. For more information about using, configuring, and managing the resolver cache, see “Resolver caching” on page 750.

Forwarders

Normally, name servers answer queries from cached data or, if that does not succeed, they attempt to contact other name servers identified in their data files as authoritative for certain domains. However, name servers can also be configured to contact special servers called *forwarders* before contacting the name servers listed in their data files. If a forwarder cannot process the query and if the local name server is not a forward-only name server, the local name server contacts the name servers in its data files. A forward-only name server relies completely on its forwarders. It does not try to contact other servers to find out information if the forwarders do not give it an answer.

The forwarding function is useful for reducing the number of queries to servers on the Internet and for creating a large cache of information on forwarders. It is also a useful function for providing Internet access for local servers that, for one reason or another, do not have access themselves.

Stealth server

A *stealth server* is a server that answers authoritatively for a zone, but is not listed in that zone's NS records. Stealth servers can be used as a way to centralize distribution of a zone, without having to edit the zone on a remote name server. When the master file for a zone is on a stealth server in this way, it is often referred to as a *hidden primary* configuration. Stealth servers can also be a way to keep a local copy of a zone for rapid access to the zone's records, even if all official name servers for the zone are inaccessible.

Resolvers

Programs that query a name server are called *resolvers*. Because many TCP/IP applications need to query the name server, a set of routines is usually provided for application programmers to perform queries. On z/OS, these routines are available in the resolver provided by z/OS Communications Server.

z/OS Communications Server provides programs for interactively querying a name server:

- NSLOOKUP (TSO)
- **onslookup**/**nslookup** (z/OS UNIX)
- DIG (TSO)
- **dig** (z/OS UNIX)
- **host**

Note: The **nsupdate** program also makes queries to name servers as part of its operations.

For information on these programs, see z/OS Communications Server: IP System Administrator's Commands.

The **onslookup** command and the **dig** command use the resolver initialization facilities of the resolver provided by z/OS Communications Server, but use their own resolver for any additional resolver facilities needed.

Resolver directives for nslookup

The onslookup program uses the following resolver directives (TCPIP.DATA statements):

- domain/domainorigin
- search
- nameserver/nsinteraddr
- options debug/options ndots

Resolver directives for dig

The dig program uses the following resolver directives (TCPIP.DATA statements):

- domain/domainorigin
- search
- nameserver/nsinteraddr
- options ndots

Query Packets

Resolvers operate by sending query packets to a name server. A query packet contains the following fields:

- Domain name
- Query type
- A query class

RFC 1035 defines the query types and query classes. The name server attempts to match the three fields of the query packet to its database. For flexibility, the following wildcard query types are defined:

Type
Description

ANY
Indicates any record type for the domain name.

AXFR
Indicates the query type used by secondary name servers to transfer all records in the zone. (The query class is set to IN when using the AXFR query type.)

MAILB
Indicates any mailbox records for the domain name.

The name server can return the following query responses:

Response
Description

Authoritative
Is returned from a primary or secondary name server. The name server contains all the domain data used to define the zone for the specified query.

BADVERS
The name server received a request that contained a EDNS0 version that was not valid.

Nonauthoritative
Is returned from a cache kept by a name server. The cache does not contain the domain data used to define the zone for the specified query.

Format Error
The name server found an error in the query packet sent by the resolver.

Name Error
No resource records of any type (including wildcards) exist for the domain name specified.

NXDOMAIN (negative)
No records of the requested type were found for the domain name specified.

Not-implemented
The name server does not support the type of query requested.

NOTAUTH
The name server is not authoritative for the zone.

NOTZONE
A dynamic update failed because the name to be updated is not contained within the given zone.

NXRRSET
A dynamic update failed because the prerequisites were not satisfied. The Resource Record set existed when the prerequiste stated it should not.

Referral
Contains the addresses of other name servers that might be able to answer the query. A referral response is returned when a recursive query is not supported, not requested, or cannot be answered because of network connectivity.

Refused
The name server refuses to perform the specified operation. For example, some root name servers limit zone transfers to a set number of IP addresses.

YXDOMAIN
DNAME mapping failed because the new name was too long.

YXRRSET
A dynamic update failed because the prerequisites were not satisfied. The Resource Record set did not exist when the prerequiste stated it should.

Resource Records

Data from a name server is stored and distributed in a format known as a resource record. Each response from a name server can contain several resource records, which can contain a variety of information. The format of a response is defined in RFC 1035. It includes the following sections:

- A question section, echoing the query for which the response is returned.
- An answer section, containing resource records matching the query.
- An additional section, containing resource records that do not match the query, but might provide useful information for the client. For example, the response to a query for the host name of a name server for a specific zone includes the IP address of that name server in the additional section.
- An authority section, containing information specific to the type of response made to the query. If a referral is returned, this section contains the domain names of name servers that could provide an authoritative answer. If a negative response is returned indicating the name does not exist, this section contains a Start Of Authority (SOA) record defining the zone of authority of the responding name server.

Querying name servers

This topic describes how to use the nslookup command to query the name server. onslookup is an alias of nslookup in the z/OS UNIX environment.

Notes:

1. The z/OS UNIX nslookup command runs only from the z/OS shell. The nslookup command can query the name server from TSO or the z/OS shell. However, only the legacy TSO version of NSLOOKUP is available from TSO. See z/OS Communications Server: IP System Administrator's Commands for detailed information.
2. The **host** and **dig** commands are another way to query name servers from the z/OS shell. For information on the **host** and **dig** commands, see z/OS Communications Server: IP System Administrator's Commands.

nslookup command

The z/OS UNIX nslookup and TSO NSLOOKUP commands can be used to query the name server to perform the following tasks:

- Identifying the location of name servers
- Examining the contents of a name server database
- Establishing the accessibility of name servers

Note: sortlist is not supported by nslookup

The z/OS UNIX nslookup and TSO NSLOOKUP commands have two modes of operation: interactive mode and command mode. The address of the default name server comes from the resolver configuration data. In the sample data below, the default domain is raleigh.ibm.com, and the default name server is at 9.37.34.149. If that name server fails to respond, the one at 9.37.34.7 is used.

```
domain   raleigh.ibm.com
nameserver   9.37.34.149
nameserver   9.37.34.7
```

Entering the interactive mode

Interactive mode can be used to repetitively query one or more name servers for information about various hosts and domains, to display that information on the console, and, in some cases, to write response data to a file.

You can enter the interactive mode under the following conditions only:

- No arguments are supplied on command invocation or the -v option is specified; the default name server is used.
- The first argument is a hyphen, and the second argument is the host name or Internet address of a name server.

For a complete description of the z/OS UNIX and TSO nslookup interactive modes, see z/OS Communications Server: IP System Administrator's Commands.

Entering the command line mode

The command line mode displays or stores the output from the query supplied as part of the invocation string and then exits.

To enter the command line mode, provide a complete query with the z/OS UNIX nslookup command invocation string.

For a complete description of the z/OS UNIX and TSO nslookup command line modes, see z/OS Communications Server: IP System Administrator's Commands.

nslookup configuration

There are only two places to specify **nslookup** options: as command options, or from the resolver configuration data set. Only a few of the options may be set in the resolver configuration data set. The command options always have precedence over any option configured in the resolver configuration data set.

Only the following options can be specified in the resolver configuration data set for v9 **nslookup**:

- nameserver/nsinteraddr
- options ndots: *n*
- search
- domain/domainorigin

The z/OS UNIX **nslookup** command uses a private resolver that is different from the resolver used by other z/OS UNIX socket programs. The z/OS UNIX **nslookup** command has the following functional differences:

- z/OS UNIX **nslookup** does not use SiteTables (for example, /etc/hosts) for host name resolution.
- Only the built-in translation table is used for **nslookup**.

For a complete discussion of resolver configuration files, see Chapter 2, “IP configuration overview,” on page 11.

Recommended reading

The latest edition of *DNS and BIND* by Paul Albitz and Cricket Liu (O'Reilly & Associates, Inc.) gives a comprehensive description of DNS and BIND.

You can subscribe to a BIND users mailing list at http://lists.isc.org/mailman/listinfo.

DNS protocols are described in various Request for Comments (RFC) papers and Internet drafts. RFCs outline existing protocols, suggest new protocols, and establish standards for the Internet protocol suite. Internet drafts are proposals, techniques, and mechanisms that document Internet Engineering Task Force (IETF) work-in-progress.

For information about obtaining RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

Resolver API calls

Application programs invoke resolver functions by using resolver API calls such as gethostbyname() and getaddrinfo(). The z/OS resolver supports a number of IBM APIs, although not all APIs support all possible resolver API calls. The z/OS resolver is invoked by the following resolver API calls:

- Application programs that use the gethostbyaddr() and gethostbyname() resolver calls from the following IBM APIs:
  - z/OS XL C/C++ runtime library functions in z/OS XL C/C++ Runtime Library Reference
  - z/OS UNIX System Services (z/OS UNIX) callable services in z/OS UNIX System Services Programming: Assembler Callable Services Reference
  - z/OS Communications Server C/C++ API, z/OS Communications Server Callable and Macro API, z/OS Communications Server REXX API, and z/OS Communications Server PASCAL API in z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference
- Application programs that use the getaddrinfo(), getnameinfo(), and freeaddrinfo() resolver calls from the following IBM APIs:
  - z/OS XL C/C++ runtime library functions in z/OS XL C/C++ Runtime Library Reference
  - z/OS UNIX System Services (z/OS UNIX) callable services in z/OS UNIX System Services Programming: Assembler Callable Services Reference
  - z/OS Communications Server Callable and Macro API and z/OS Communications Server REXX API in z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference
- Application programs that use the sethostent(), gethostent(), and endhostent() resolver calls from the following IBM APIs:
  - z/OS XL C/C++ runtime library functions in z/OS XL C/C++ Runtime Library Reference
  - z/OS Communications Server C/C++ API in z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference

T he **dig**, **nslookup**, and **nsupdate** z/OS UNIX commands provide their own unique resolver services. When their resolver initializes, it uses the appropriate TCPIP.DATA information, including information from the global TCPIP.DATA file, if one is specified. For more information, see "The resolver and the global TCPIP.DATA file" on page 741.

Starting the resolver

Prior to starting the resolver, you must have a user ID defined for the resolver started procedure that includes an OMVS segment, so that the resolver can use z/OS UNIX services. For information about defining and assigning a user ID for started procedures, see z/OS Security Server RACF Security Administrator's Guide.

Requirement: If the resolver user ID does not have an OMVS segment and you are not using the automatic assignment of unique UNIX identities support to automatically generate OMVS segments, you must explicitly configure an OMVS segment for the resolver user ID. For more information about RACF and z/OS UNIX, see z/OS Security Server RACF Security Administrator's Guide.

Guideline: The HOME directory in the OMVS segment for the resolver user ID must be accessible when the resolver is started. Specify the root directory [HOME(/)] to avoid problems.

There are two ways that the resolver can be started:

- Use z/OS UNIX to start the resolver

  The resolver is started when z/OS UNIX is initialized. This method of starting the resolver ensures that applications that require resolver services are not started before the resolver starts. If you use z/OS UNIX to start the resolver, you can use the default resolver settings (see "The default resolver settings" on page 738), or you can optionally create a start procedure and define the address space (see "Customizing the resolver" on page 738).

- Use automation tools to start the resolver

  The resolver is started by issuing the MVS START operator command. You must customize the resolver to use this starting method; see "Customizing the resolver" on page 738 for more information. If you use this method of starting the resolver, then it is possible that an application that needs resolver services (such as INETD) is started before the resolver address space is initialized. In this case, you might need to remove the starting of INETD from the z/OS UNIX /etc/rc file and start INETD with automation after the resolver has initialized.

If you are using z/OS UNIX to start the resolver, then the following actions occur when z/OS UNIX is initialized:

- z/OS UNIX uses SUB=MSTR to start the resolver (the resolver does not require JES); for information about SUB=MSTR, see z/OS MVS JCL Reference
- z/OS UNIX issues an informational message that contains the name of the procedure that it is starting:

  `BPXF224I THE RESOLVER_PROC, procname, IS BEING STARTED`

  Rule: If the RESOLVER_PROC statement is not present or is specified with the procedure name DEFAULT, the *procname* value is RESOLVER, even though a start procedure was not used. For more information about the RESOLVER_PROC statement, see z/OS MVS Initialization and Tuning Reference.
- If the start procedure is not found or if it contains a JCL error, then error messages for the z/OS START command are issued and the resolver address space does not initialize.
- If the resolver address space cannot be started, z/OS UNIX initialization continues.

The default resolver settings

You do not have to perform any customization steps to use the resolver. The default resolver starts automatically when z/OS UNIX is initialized; you cannot use automation tools to start the resolver unless you customize it. z/OS UNIX starts a resolver address space, which uses the assigned name RESOLVER, using the system default procedure IEESYSAS. The resolver uses the applicable native MVS or z/OS UNIX search order to find TCPIP.DATA statements, without a GLOBALTCPIPDATA or a DEFAULTTCPIPDATA specification.

The following resolver functions are active by default:

- System-wide caching, which uses the default maximum cache size, the default maximum time-to-live (TTL) value, and the default setting for the cache reordering function. See "Resolver caching" on page 750.
- Network operator notification of unresponsive Domain Name System (DNS) servers, using the default threshold setting. See "Monitoring the responsiveness of Domain Name System name servers" on page 761.

The Extension Mechanisms for DNS (EDNS0) standards function is always active. For more information, see "Extension Mechanisms for DNS standards and the resolver" on page 770.

Customizing the resolver

You can customize resolver functions, or enable or disable certain resolver functions, by using resolver configuration statements in a resolver setup file. For example, you can control which resolver statements are used by all applications or TCP/IP stacks for name resolution by specifying the GLOBALTCPIPDATA statement (see "The resolver setup file" on page 738 for more information about the statements that are supported by the resolver setup file). If you want to customize resolver functions, you must create a resolver setup file and define the resolver address space.

The resolver setup file

The resolver setup file is an optional file (either an MVS data set or a z/OS UNIX file) that contains resolver configuration statements that you can use to customize resolver functions.

If the resolver setup file is an MVS data set, the file must have the following characteristics:

- Use sequential (PS) or partitioned (PO) organization
- Use fixed (F) or fixed block (FB) format
- Contain a logical record length (LRECL) that is 80 - 256 bytes in length
- Contain any valid blocksize (BLKSIZE) for a fixed block
- If you think you will need to modify the setup file, use a member of an MVS partitioned data set

If the resolver setup file is a z/OS UNIX file, the file can be located in any directory. The maximum line length that is supported is 256 characters. If a line is longer than 256 characters, then the line is truncated to 256 characters before it is processed.

If you do not use a resolver setup file, the resolver uses the applicable native MVS or z/OS UNIX search order without any additional information and performs the following functions using the default settings:

- Enables system-wide caching (using the default maximum cache size and the default time-to-live [TTL] value)
- Monitors Domain Name System (DNS) name server responsiveness (using the default threshold setting)

The following statements are supported by the resolver setup file:

- Comments (; or #)
- CACHE

  Enables system-wide caching of DNS queries that have been resolved. System-wide caching is enabled by default, but you can explicitly enable it using this statement. For more information about caching, see “Resolver caching” on page 750.
- CACHEREORDER

  Enables system-wide reordering of a cached list of IP addresses in response to resolver API queries for the associated host name. If you do not specify this statement, system-wide reordering of a cached list is disabled by default.
- CACHESIZE

  Defines the amount of storage that can be allocated by the resolver to manage cached records.
- COMMONSEARCH

  Indicates that the same search order for local host files is used for both IPv4 and IPv6 name queries.
- DEFAULTIPNODES

  Identifies the default local host file.
- DEFAULTTCPIPDATA

  Identifies a default TCPIP.DATA file. The file that is specified by the DEFAULTTCPIPDATA statement becomes the last file that is searched by the resolver for resolver configuration information. If you do not specify the DEFAULTTCPIPDATA statement, the default file is TCPIP.TCPIP.DATA. For more information, see “The resolver and the global TCPIP.DATA file” on page 741.
- GLOBALIPNODES

  Identifies a local host file that contains hard-coded IP addresses and host names that can be used globally.
- GLOBALTCPIPDATA

  Identifies the file that is the first file that is searched by the resolver for resolver configuration information. Parameters that you specify in the file that is identified by the GLOBALTCPIPDATA statement become the global settings for the entire MVS image and for all TCP/IP stacks. For more information, see “The resolver and the global TCPIP.DATA file” on page 741.

  Restriction: You must code the GLOBALTCPIPDATA statement if the AUTOQUIESCE operand is coded on the UNRESPONSIVETHRESHOLD statement. If you cannot ensure that all DNS IP addresses are accessible from all your TCPIP stacks, you should not use a global TCPIP.DATA file, and you should

not code AUTOQUIESCE on the UNRESPONSIVETHRESHOLD setup statement in your resolver setup file. For more information about the UNRESPONSIVETHRESHOLD statement, see z/OS Communications Server: IP Configuration Reference.

- MAXNEGTTL

  Defines the maximum amount of time that the resolver can use negative cache resource information that it receives from a Domain Name System (DNS) name server.

- MAXTTL

  Defines the amount of time that the resolver can use resource information that it receives from a Domain Name System (DNS) name server.

- NOCACHE

  Disables system-wide caching of DNS response data. If you do not specify this statement, system-wide caching is enabled by default.

- NOCACHEREORDER

  Disables system-wide reordering of a cached list of IP addresses in response to resolver API queries for the associated host name. System-wide reordering of a cached list of IP addresses is disabled by default, but you can explicitly disable it by using this statement.

- NOCOMMONSEARCH

  Indicates that a different search order for local host files is used for IPv4 and IPv6 name queries.

- UNRESPONSIVETHRESHOLD

  Identifies the threshold value that determines when the resolver declares a DNS name server to be unresponsive. If the percentage of query failures to a name server during a fixed interval is greater than or equal to this threshold value, the resolver considers the name server to be unresponsive. The resolver might stop sending DNS queries that are generated by an application to unresponsive name servers, depending on the values that are coded on this statement. For more information, see “Monitoring the responsiveness of Domain Name System name servers” on page 761.

  Restriction: You must code the GLOBALTCPIPDATA statement if the AUTOQUIESCE operand is coded on the UNRESPONSIVETHRESHOLD statement. If you cannot ensure that all DNS IP addresses are accessible from all your TCPIP stacks, you should not use a global TCPIP.DATA file, and you should not code AUTOQUIESCE on the UNRESPONSIVETHRESHOLD setup statement in your resolver setup file. For more information about the UNRESPONSIVETHRESHOLD statement, see z/OS Communications Server: IP Configuration Reference.

For more information about resolver setup statements, see z/OS Communications Server: IP Configuration Reference.

Note: The name of the resolver setup file most recently used to configure the resolver is displayed as part of the MODIFY RESOLVER command output. See z/OS Communications Server: IP System Administrator's Commands for more details.

Resolver processing of the setup file when the resolver is started

During resolver address space initialization, the resolver parses the entire resolver setup file using the following rules:

- When a setup statement has a syntax error, the resolver generates a warning message to the console and the setup statement is skipped.
- When the resolver does not recognize a setup statement, the resolver generates a warning message to the console and the setup statement is skipped.
- When the resolver parses multiple instances of the same setup statement that have no syntax errors, the resolver uses the setting on the last instance that is parsed.
- When the resolver completes parsing of the resolver setup file and did not parse any instances of a resolver setup statement that are syntactically correct, the resolver takes the default value for that setup statement

- When a resolver setup statement (for instance, GLOBALTCPIPDATA) specifies an MVS data set name or a z/OS UNIX file name that the resolver cannot open or access, the resolver considers that to be a syntax error.
- When there are multiple instances of a setup statement that have no syntax errors and that specify an MVS data set name or z/OS UNIX file name, the resolver attempts to open or access the data set or file specified on only the last instance that is parsed. If the resolver cannot open or access this data set or file, the resolver proceeds as if this setup statement had not been specified.
- When the resolver cannot open the resolver setup file, the resolver uses all the default settings for the configuration.
- When the resolver setup file is the wrong format, the resolver uses all the default settings for the configuration.
- If the resolver generates any warning messages due to syntax errors or unrecognized statements during resolver address space initialization, the resolver issues message EZD2038I at the conclusion of resolver address space initialization.
- If the resolver does not generate any warning messages during resolver address space initialization, the resolver issues message EZZ9291I when the address space is initialized.

Results:

- Consider a resolver setup file that contains the following statements:

```
CACHE
NOCACHE
CACH
```

  The resolver issues a warning message for the line specifying the value CACH and ignores the value. Because NOCACHE was parsed after the valid statement CACHE, the resolver uses the NOCACHE setup statement and caching is disabled for this system.

- Consider a resolver setup file that contains the following statements:

```
GLOBALTCPIPDATA(/etc/tcpip.data)
GLOBALTCPIPDATA(/etc/newer.tcpip.data)
DEFAULTTCPIPDATA(/etc/default.tcpip.data)
```

  Each of these statements is syntactically correct, so the resolver uses the value coded on the last GLOBALTCPIPDATA statement as the name of the global TCPIP.DATA file. The resolver attempts to access the files /etc/newer.tcpip.data and /etc/default.tcpip.data. If the file /etc/newer.tcpip.data does not exist, the resolver generates a warning message and proceeds as if no global file were specified; the resolver does not attempt to use /etc/tcpip.data as the global TCPIP.DATA file. If /etc/default.tcpip.data does not exist, the resolver generates a warning message and proceeds as if no default file were specified.

The resolver and the global TCPIP.DATA file

The optional GLOBALTCPIPDATA statement identifies a global TCPIP.DATA file, in which you can specify global settings. If you specify a global TCPIP.DATA file, you can control which resolver statements are used for name resolution and you do not need to merge resolver statements from multiple files. TCPIP.DATA statements in this file are the first that are searched, regardless of which socket API library you are using.

If you use a global TCPIP.DATA file, you can specify the following resolver statements, or you can use the default values. These statements are required by the resolver to process queries.

- DomainOrigin or Domain
- NSInterAddr or NameServer
- NSPortAddr
- ResolveVia
- ResolverTimeOut

- ResolverUDPRetries
- Search
- SortList

If you specify any of these statements in the global TCPIP.DATA file, those settings become the global settings for this MVS image and for all users of resolver services, across the entire system. If you do not specify one of these statements in the global TCPIP.DATA file, the resolver uses the default values. The search continues beyond the global TCPIP.DATA file, but any of these resolver statements that are specified in files that are located lower in the search order are ignored.

If you do not identify a global TCPIP.DATA file, then the resolver uses the regular search order until it finds a local TCPIP.DATA file.

Restriction: You must code the GLOBALTCPIPDATA statement if you want to use the autonomic quiescing of unresponsive name servers function. For more information, see "Monitoring the responsiveness of Domain Name System name servers" on page 761.

If you do not specify one of these statements in this local TCPIP.DATA file, the resolver uses the default values. The setting that the resolver uses applies only to this application, not to all users of resolver services. In any case, the search order depends on whether the native MVS or z/OS UNIX application environment is in use. The search order for the local hosts table (HOSTS.*xxxx*INFO, ETC.IPNODES, /etc/hosts, or /etc/ipnodes) remains the same.

You can specify other statements in the global TCPIP.DATA file (for information about configuration statements in TCPIP.DATA, see z/OS Communications Server: IP Configuration Reference). There are some statements, like Trace Resolver, for which you likely do not want a global setting. If you do not specify one or more of these other statements in the global file, the resolver uses the regular search order until it finds a local TCPIP.DATA file. If you do not specify these other statements in the local TCPIP.DATA file, the resolver uses the default values.You can implement these global settings gradually, in case there are private TCPIP.DATA files that are in use on your system about which you are unaware.

You can identify a global TCPIP.DATA file in a CINET (common INET) environment. You specify in the global TCPIP.DATA file all the resolver statements for which you want to set a global value, and the resolver conducts its searches in the same way as in a non-CINET environment. However, if you do use a global TCPIP.DATA file in a CINET environment, all the resolver statements must be usable by all stacks. For example, the IP addresses that are specified by the NameServer statement must be accessible from all stacks. If the IP addresses are not accessible, then you should not use a global TCPIP.DATA file; use multiple TCPIP.DATA data sets instead. If you use multiple TCPIP.DATA data sets, you should not code the AUTOQUIESCE operand on the UNRESPONSIVETHRESHOLD statement in the resolver setup file.

You can use the DEFAULTTCPIPDATA setup statement to specify a TCPIP.DATA file as the last TCPIP.DATA file that is searched, instead of TCPIP.TCPIP.DATA. You can specify any file as the default file.

Steps for creating a resolver setup file

The SETUP DD statement in the start procedure for the resolver points to the resolver setup file. The setup file can be an MVS data set or a z/OS UNIX file.

Procedure

Perform the following steps to create a resolver setup file:

1. Use an MVS data set or a z/OS UNIX file for your setup file, depending on your requirements.
2. Customize the search order that the resolver uses to resolve queries by specifying one or more of the following statements:
   - Specify the GLOBALTCPIPDATA statement to identify the global TCPIP.DATA file that becomes the first file that is searched. Parameters that you specify in this file become the global settings for the entire MVS image and for all users of resolver services, across the entire system.

     Restriction: You must code the GLOBALTCPIPDATA statement if the AUTOQUIESCE operand is coded on the UNRESPONSIVETHRESHOLD statement. If you cannot ensure that all DNS IP

addresses are accessible from all your TCPIP stacks, you should not use a global TCPIP.DATA file, and you should not code AUTOQUIESCE on the UNRESPONSIVETHRESHOLD setup statement in your resolver setup file. For more information about the UNRESPONSIVETHRESHOLD statement, see z/OS Communications Server: IP Configuration Reference.

- Specify the DEFAULTTCPIPDATA statement to identify a default TCPIP.DATA file. The file specified by the DEFAULTTCPIPDATA statement becomes the last file that can be searched. If you do not specify the DEFAULTTCPIPDATA statement, the default file is TCPIP.TCPIP.DATA.
- Specify either the GLOBALIPNODES statement or the DEFAULTIPNODES statement to identify a local host file. The GLOBALIPNODES statement identifies a local host file that contains hard-coded IP addresses and host names that can be used globally. The DEFAULTIPNODES statement identifies the default local host file.

  Result: You can specify both the GLOBALIPNODES statement and the DEFAULTIPNODES statement, but the resolver search order for local host files ensures that only the global IPNODES file is used. For more information, see "Search orders used in the z/OS UNIX environment" on page 777 and "Search orders used in the native MVS environment" on page 781.
- Specify the COMMONSEARCH statement or the NOCOMMONSEARCH statement. The COMMONSEARCH statement indicates that the same search order for local host files is used for both IPv4 and IPv6 name queries, and for MVS and UNIX searches. The NOCOMMONSEARCH statement indicates that different search orders for local host files are used for IPv4 and IPv6 name queries, and for MVS and UNIX searches. For more information, see "Search orders used in the z/OS UNIX environment" on page 777 and "Search orders used in the native MVS environment" on page 781.
- Specify the CACHE statement or the NOCACHE statement. The CACHE statement enables system-wide caching of Domain Name System (DNS) queries that have been resolved. System-wide caching is enabled by default, but you can explicitly enable it using this statement. See "Resolver caching" on page 750 for more information about caching. The NOCACHE statement indicates that you do not want to cache DNS response data.
  - If you are using system-wide caching, specify the CACHESIZE statement to define the amount of storage that can be allocated by the resolver to manage cached records. If you specify this statement and the NOCACHE statement, the CACHESIZE statement is ignored.
  - If you are using system-wide caching, specify the MAXNEGTTL statement to define the amount of time that the resolver can use negative cache resource information that it receives from a Domain Name System (DNS) name server. Negative cache information represents DNS resource lookups that result in NXDOMAIN responses, and indicate that the resource name (hostname or IP address) or the DNS zone does not exist. If you specify this statement and the NOCACHE statement, the MAXNEGTTL statement is ignored.
  - If you are using system-wide caching, specify the MAXTTL statement to define the amount of time that the resolver can use resource information that it receives from a name server. If you specify this statement and the NOCACHE statement, the MAXTTL statement is ignored.
  - If you are using system-wide caching, you can enable the resolver to reorder the cached list of IP addresses when the resolver responds to a host name resolution request.
    - To enable reordering of a list of cached IP addresses in a round-robin manner, specify the CACHEREORDER statement.
    - To disable reordering of a list of cached IP addresses, specify the NOCACHEREORDER statement.

    If you specify the NOCACHE statement, the CACHEREORDER and NOCACHEREORDER statements are ignored.
- Specify the UNRESPONSIVETHRESHOLD statement to define the threshold value that is used by the resolver to declare that a DNS name server is unresponsive. The monitoring of unresponsive name servers is enabled by default, but you can explicitly enable it using the UNRESPONSIVETHRESHOLD statement. In addition to specifying the threshold value, use the UNRESPONSIVETHRESHOLD statement to indicate how you want the resolver to react when an unresponsive name server is detected:

- Specify only the threshold value on the UNRESPONSIVETHRESHOLD statement to indicate that the resolver should only alert the operator that a name server is unresponsive. The resolver will continue to send DNS queries that are generated by an application to the name server.
- Specify the AUTOQUIESCE operand on the UNRESPONSIVETHRESHOLD statement to indicate that the resolver should stop sending DNS queries that are generated by an application to unresponsive name servers.

  Restriction: You must code the GLOBALTCPIPDATA statement if the AUTOQUIESCE operand is coded on the UNRESPONSIVETHRESHOLD statement. If you cannot ensure that all DNS IP addresses are accessible from all your TCPIP stacks, you should not use a global TCPIP.DATA file, and you should not code AUTOQUIESCE on the UNRESPONSIVETHRESHOLD setup statement in your resolver setup file. For more information about the UNRESPONSIVETHRESHOLD statement, see z/OS Communications Server: IP Configuration Reference.

If you do not want the resolver to maintain awareness of unresponsive name servers, you can use the UNRESPONSIVETHRESHOLD statement to disable resolver monitoring of name server responsiveness to DNS queries.

Results

The following example setup file is located in SEZAINST as member EZBRECNF (alias RESSETUP):

```
;
;   IBM Communications Server for z/OS
;   SMP/E distribution name: EZBRECNF
;
;   5694-A01 Copyright IBM Corp. 2002, 2019,
;   Licensed Materials - Property of IBM
;
;   Function: Sample Resolver setup file
;
;
;  The following statement defines the final search location for
;   TCPIP.DATA statements. It will replace TCPIP.TCPIP.DATA
;  It may be an MVS data set or HFS file.
;
DEFAULTTCPIPDATA('TCPIP.TCPIP.DATA')
;
#  The following statement defines the first search location for
#   TCPIP.DATA statements. It may be an MVS data set or HFS file.
;
;  Update with the correct data set or HFS file name
;
; GLOBALTCPIPDATA('TCPCS.SYS.TCPPARMS(GLOBAL)')
;
; GLOBALTCPIPDATA(/etc/tcpipglobal.data)
;
#  The following statement defines the first search location for
#   IPNODES statements. It may be an MVS data set or HFS file.
;
;  Update with the correct data set or HFS file name
;
; GLOBALIPNODES('TCPCS.SYS.TCPPARMS(IPNODES)')
;
; GLOBALIPNODES('TCPCS.ETC.IPNODES')
;
; GLOBALIPNODES(/etc/ipnodes)
;
#  The following statement defines the final search location for
#   IPNODES statements. It may be an MVS data set or HFS file.
;
;  Update with the correct data set or HFS file name
;
; DEFAULTIPNODES('TCPCS.SYS.TCPPARMS(IPNODES)')
;
; DEFAULTIPNODES('TCPCS.ETC.IPNODES')
;
; DEFAULTIPNODES(/etc/ipnodes)
;
#  The following statement defines if the common search order
#   should be used or not.
;
```

```
NOCOMMONSEARCH
;
; COMMONSEARCH
;
#  The following statement defines if system-wide resolver caching
#   should be used or not.
;
CACHE
;
; NOCACHE
;
#  The following statement defines if the resolver reorders
#   the list of cached IP addresses in response to resolution
#   requests for the associated host name.
;
NOCACHEREORDER
;
; CACHEREORDER
;
#  The following statement defines the amount of storage that the
#   resolver can use for holding system-wide resolver cache data.
;
CACHESIZE(200M)
;
#  The following statement defines the maximum amount of time that
#   the resolver can use resource information that was cached as
#   a result of a query to a name server.
;
MAXTTL(2147483647)
;
#  The following statement defines the maximum amount of time that
#   the resolver can use information about non-existent resources
#   that was cached as a result of a query to a name server.  If not
#   specified the statement defaults to the value specified in MAXTTL.
;
MAXNEGTTL(2147483647)
;

#  The following statement defines the threshold value for declaring
#   a name server as unresponsive.  The AUTOQUIESCE option defines
#   whether resolver temporarily stops using unresponsive name servers.
#   You must code the GLOBALTCPIPDATA statement when you specify the
#   AUTOQUIESCE option on the UNRESPONSIVETHRESHOLD statement.
;
UNRESPONSIVETHRESHOLD(25)
;
; UNRESPONSIVETHRESHOLD(100,AUTOQUIESCE)
```

The resolver address space

The resolver address space must be started before any application or TCP/IP stack resolver calls can occur. When the address space starts, it reads an optional resolver setup data set that is pointed to by the SETUP DD card in the resolver JCL procedure. To use the functions that are provided by the GLOBALTCPIPDATA statement and other statements, you must define a resolver address space. You use a BPXPRMxx statement, RESOLVER_PROC, to specify the procedure name, if any, to be used by z/OS UNIX to start the resolver address space. If the RESOLVER_PROC statement is not in the BPXPRMxx parmlib member or is specified with the procedure name DEFAULT, z/OS UNIX starts a resolver address space that has the assigned name RESOLVER.

Steps for defining the resolver address space

You must define the resolver address space to use the functions provided by the GLOBALTCPIPDATA statement and other statements.

Before you begin

You must have already created a resolver setup file, if you are using one.

Procedure

Perform the following steps to define the resolver address space:

1. Create a start procedure.

The start procedure has the following requirements:

- The procedure must not contain any DD cards that specify SYSOUT=*.
- The procedure must be in a data set that is specified by the IEFPDSI DD card specification of the MSTJCLxx PARMLIB member. If the procedure is not in this location, the resolver will not start. For information about MSTJCL, see z/OS MVS Initialization and Tuning Reference.

To process its definitions, the resolver might need to allocate data sets or files. For those definitions, such as GLOBALTCPIPDATA, DEFAULTTCPIPDATA, /etc/hosts, HOSTS.SITEINFO, and HOSTS.ADDRINFO, allocation messages appear in the JES joblog. For long-running applications that heavily use resolver services, such as IBM Tivoli NetView for z/OS, consider using a started job that specifies MSGLEVEL=(1,0) to eliminate all allocation messages. This specification could also eliminate allocation messages that might be useful for problem analysis. For information about started jobs and the MSGLEVEL parameter, see z/OS MVS JCL Reference.

See the example resolver startup procedure.

2. In the SETUP DD JCL statement, specify the location of the setup file.
3. Grant read access (using RACF or other security program) to the following files for the user ID that is assigned to the resolver address space:
   - SYS1.PARMLIB
   - The resolver setup file
   - The file specified by the GLOBALTCPIPDATA statement, if you are using one
   - The file specified by the DEFAULTTCPIPDATA statement, if you are using one
   - The file specified by the GLOBALIPNODES statement, if you are using one
   - The file specified by the DEFAULTIPNODES statement, if you are using one
4. If you are using any of the following files, grant read access to the files for the user IDs or jobs that are using TCP/IP facilities:
   - The file specified by the GLOBALTCPIPDATA statement
   - The file specified by the DEFAULTTCPIPDATA statement
   - The file specified by the GLOBALIPNODES statement
   - The file specified by the DEFAULTIPNODES statement
   - /etc/hosts
   - /etc/ipnodes
   - /etc/services
   - HOSTS.SITEINFO
   - HOSTS.ADDRINFO
   - ETC.IPNODES

   If you do not specify the correct permission bit settings for a file to allow that file to be read, error message IEC141I 013-C0 is issued. Other error messages might also be issued that indicate that a file cannot be read.
5. (Optional) If you want z/OS UNIX to start the resolver (rather than using automation to start the resolver with the MVS START operator command), specify the resolver start procedure name as the *procname* value in the RESOLVER_PROC(*procname*) statement of the BPXPRMxx parmlib member.

   See “Starting the resolver” on page 737 for information about starting the resolver.

   Guideline: The default procedure name is RESOLVER. If you want to specify a procedure that uses the name RESOLVER, then specify RESOLVER for the *procname* value.

   If you do not specify the RESOLVER_PROC statement or if you specify DEFAULT, then z/OS UNIX starts a resolver address space using the system default procedure IEESYSAS with the assigned name RESOLVER. If you do not want to use z/OS UNIX to start the resolver, you must use the MVS START command to start the resolver address space.

6. If you use automation to start the resolver, consider looking for message EZD2038I and taking corrective action based on the warning messages that are generated by the resolver.

   The resolver issues messages during resolver address space initialization identifying the errors that exist in the resolver setup file. If the resolver issues message EZD2038I at the end of resolver address space initialization, there were setup statements with syntax errors or setup statements that were not recognized in the setup file. The errors might include errors accessing or opening one or more of the files specified on resolver setup statements, such as the GLOBALTCPIPDATA statement.

Results

The following example resolver start procedure is located in SEZAINST as member EZBREPRC (alias RESOPROC).

```
//RESOLVER PROC PARMS='CTRACE(CTIRES00)'
//*
//*   IBM Communications Server for z/OS
//*   SMP/E distribution name: EZBREPRC
//*
//*   5694-A01 Copyright IBM Corp. 2001, 2010.
//*   Licensed Materials - Property of IBM
//*
//*   Function: Start Resolver
//*
//EZBREINI EXEC PGM=EZBREINI,REGION=0M,TIME=1440,PARM=&PARMS
//*
//*   When the Resolver is started by UNIX System Services it is
//*     started with SUB=MSTR.
//*   This means that JES services are not available to the Resolver
//*     address space. Therefore, no DD cards with SYSOUT can be used.
//*     See the MVS JCL Reference manual for SUB=MSTR considerations in
//*     section "Running a Started Task Under the Master Subsystem".
//*   This Resolver start procedure will need to reside in a data
//*     set that is specified by the MSTJCLxx PARMLIB member's
//*     IEFPDSI DD card specification. If not, the procedure will
//*     not be found and the Resolver will not start.
//*     See the MVS Initialization and Tuning Reference manual for
//*     MSTJCL considerations in section "Understanding the Master
//*     Scheduler Job Control Language"
//*
//*   SETUP contains Resolver setup parameters.
//*     See the chapter "The resolver" in the
//*     IP Configuration Guide for more information. A sample of
//*     Resolver setup parameters is included in member RESSETUP
//*     of the SEZAINST data set.
//*
//*SETUP   DD   DSN=TCPIP.TCPPARMS(SETUPRES),DISP=SHR,FREE=CLOSE
//*SETUP   DD   DSN=TCPIP.SETUP.RESOLVER,DISP=SHR,FREE=CLOSE
//*SETUP   DD   PATH='/etc/setup.resolver',PATHOPTS=(ORDONLY)
```

Managing the resolver address space

The resolver start procedure name is used with the following MVS system commands to manage the resolver address space:

- Start (S)
- Stop (P)

  You should stop and restart the resolver only when a new level of the resolver code has been installed.
- Force
- Modify (F)

  Use the MODIFY command to dynamically change resolver setup statements, to update the use of TCPIP.DATA statements, or to update the use of local host and services tables.

  You can also use the MODIFY command to delete the information that the resolver has acquired about name server capabilities, and to restart the monitoring of unresponsive DNS name servers function after the resolver has stopped monitoring because of a system error.

You can use the MODIFY FLUSH command to delete all the resolver cache data.

Guideline: If you use the MODIFY command to dynamically change resolver setup statements and the resolver detects unrecognized setup statements or setup statements with syntax errors in the setup file specified on the MODIFY command, the resolver does not process the MODIFY command.

See z/OS Communications Server: IP System Administrator's Commands for more information about these commands.

You can use the following MVS system commands to control and display the status of the resolver CTRACE facilities:

- Trace CT
- Display Trace

See z/OS Communications Server: IP Diagnosis Guide for information about using CTRACE.

Steps for manually restarting the resolver

If the resolver is stopped for any reason, manually restart the resolver.

Before you begin

When you start the resolver, specify REUSASID=YES on the START command to ensure that a reusable ASID is used. If the resolver address space is stopped enough times and you do not specify REUSASID=YES, when you restart the resolver, all available ASIDs might be exhausted, which prevents the creation of a new address space on the system. If a new address space is not created on the system, an IPL is required. For more information about tuning parameters for the maximum number of ASIDs on a system, see the MAXUSER parameter in z/OS MVS Initialization and Tuning Reference.

Procedure

Perform one of the following steps to restart the resolver:

- If you have not customized the resolver (you have not created a start procedure and you have not defined the address space), then issue the following system operator command:

```
START IEESYSAS.RESOLVER,PROG=EZBREINI,SUB=MSTR,REUSASID=YES
```

- If you have customized the resolver, issue one of the following system operator commands, where *procname* is the name of the PROCLIB member that you created:

```
START procname,REUSASID=YES
START procname,SUB=MSTR,REUSASID=YES
```

Steps for applying an interim fix to the resolver

To apply IBM-supplied interim fixes to the resolver, you do not need to stop and restart the TCPIP started task or any application programs.

Before you begin

For the short duration of time that the resolver is not running, all resolver requests will fail.

Procedure

Perform the following steps to apply an interim fix to the resolver:

1. Use SMP/E to apply the interim fix.
2. If LLA (library lookaside) is running, refresh it by issuing the following operator command:

```
MODIFY LLA,REFRESH
```

3. Stop the resolver by performing one of the following steps:

- If you have not customized the resolver (you have not created a start procedure), issue the following system operator command:

```
STOP RESOLVER
```

- If you have customized the resolver, issue the following system operator command, where *procname* is the name of the PROCLIB member that you created:

```
STOP procname
```

4. Restart the resolver by performing one of the following steps:
   - If you have not customized the resolver (you have not created a start procedure), issue the following system operator command:

```
START IEESYSAS.RESOLVER,PROG=EZBREINI,SUB=MSTR,REUSASID=YES
```

   - If you have customized the resolver, issue one of the following system operator commands, where *procname* is the name of the PROCLIB member that you created:

```
START procname,REUSASID=YES
START procname,SUB=MSTR,REUSASID=YES
```

   Rule: When you manually stop and then restart the resolver, specify REUSASID=YES on the START command to ensure that a reusable ASID is used. If the resolver address space is stopped enough times and you do not specify REUSASID=YES then when you restart the resolver, all available ASIDs might be exhausted, which would prevent the creation of a new address space on the system. If a new address space is not created on the system, an IPL is required. For more information on tuning parameters for the maximum number of ASIDs on a system, see the MAXUSER parameter in z/OS MVS Initialization and Tuning Reference.

IPv6 name servers and the resolver

The z/OS resolver can communicate to a Domain Name System (DNS) name server using either IPv4 or IPv6 communications. When the resolver is operating on an IPv6-capable system, the resolver opens either an IPv6 socket or an IPv4 socket to connect to the target DNS name server, depending on the IP addresses of the potential target name servers. If both IPv6 and IPv4 addresses are specified as IP addresses for the target name servers, the resolver opens an IPv6 socket and relies on TCP/IP stack processing of IPv4-mapped IPv6 addresses to provide the appropriate communication with the name server. When the resolver is operating on an IPv4-only system, any IPv6 address specified as a target name server is ignored.

The list of name servers to be searched, as defined using the NSINTERADDR or NAMESERVER configuration statements in the TCPIP.DATA file, is stored in the res_state control block structure. Depending on the API used, applications can obtain the res_state control block, and can examine or even modify the contents of the res_state list of name servers (nsaddr_list). The nsaddr_list structure is not designed to accommodate IPv6 addresses, and changing the structure to accommodate IPv6 addresses would adversely affect existing applications. Therefore, resolver does not save IPv6 addresses into the nsaddr_list array, and returns only the IPv4 addresses coded on the NSINTERADDR or NAMESERVER configuration statements. If only IPv6 addresses are specified for name server addresses, the nsaddr_list array returns as an empty list in res_state. If your application examines the contents of the res_state control block, in particular the nsaddr_list information, and requires that at least one name server IP address be present in the list, you need to specify at least one IPv4 address on a NSINTERADDR or NAMESERVER statement in the TCPIP.DATA file used by the application. Although nsaddr_list does not reflect the IPv6 addresses specified in the TCPIP.DATA file, resolver maintains the full list internally and uses the entire list for searching and caching purposes, unless the application modifies nsaddr_list for its own purposes.

The following z/OS function do not use the resolver's ability to communicate to a name server using IPv6:

- The TSO DIG and TSO NSLOOKUP commands do not support specification of an IPv6 address as the target server IP address, either explicitly on the command or as an entry in the TCPIP.DATA file.

If you choose to use IPv6 name servers, ensure that the setting for the MAXSOCKETS parameter on the BPXPRMxx AF_INET6 NETWORK statement is sufficiently large enough to include the number of IPv6 sockets being opened by resolver. Use the following rough calculations to determine whether the MAXSOCKETS value is large enough:

1. Determine how many concurrent resolver calls typically occur on your system. Each resolver call that involves communication with a DNS name server opens 1 or 2 sockets depending on the actual resolver API being used.
2. Determine the largest number of IPv6 sockets you have open on your system at any one point in time.
3. Add (number of concurrent resolver calls * 2) and the largest number of IPv6 sockets together. Compare this number against the setting of MAXSOCKETS on the appropriate BPXPRMxx AF_INET6 NETWORK statement.
   - If the MAXSOCKETS setting is the larger value, then the setting is acceptable.
   - If the MAXSOCKETS setting is the lesser value, update the MAXSOCKETS setting to the larger number.

Resolver functions

To use the resolver functions efficiently in your environment, you need to be familiar with the following resolver functions that are active, by default, when the resolver is started:

- Resolver caching
- Monitoring responsiveness of Domain Name System name servers
- Extension Mechanisms for DNS standards

Resolver caching

In the context of resolvers, *caching* is defined as saving and storing information from resolved DNS queries so that the information can be reused. A *cache* is the area of memory where the information is kept. The primary advantage of caching is the improved performance that is obtained by the elimination of repetitive queries to the name servers.

For example, in the configuration shown in Figure 101 on page 750, a local caching-only DNS name server is defined to provide some level of resource caching.

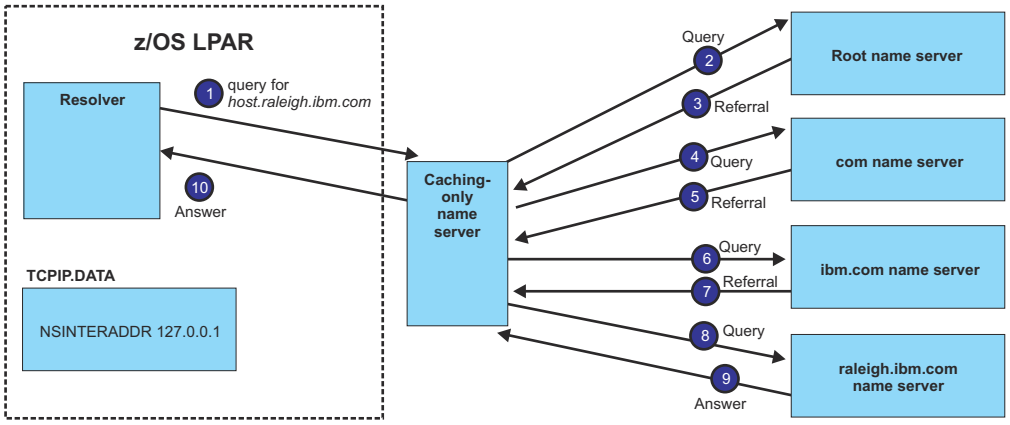


Figure 101. Local caching-only name server example

In the caching-only name server model depicted in Figure 101 on page 750, when a request for host.raleigh.ibm.com is received, the resolver contacts the local name server (see arrow 1). The local name server then must perform standard name-server processing to locate the resource, which might include contacting one or more name servers (see arrows 2 through 9). When the name server that can provide an authoritative response for the queried host name is found and the response is returned to the local name server, the local name server caches the information and forwards it back to the resolver (arrow 10).

However, in the caching-only name server model depicted in Figure 101 on page 750, subsequent requests to the resolver for information about host.raleigh.ibm.com still require that a DNS query be created and forwarded to the local name server to obtain the cached results. This DNS query can be eliminated if the resolver itself caches the information, as shown in Figure 102 on page 751.

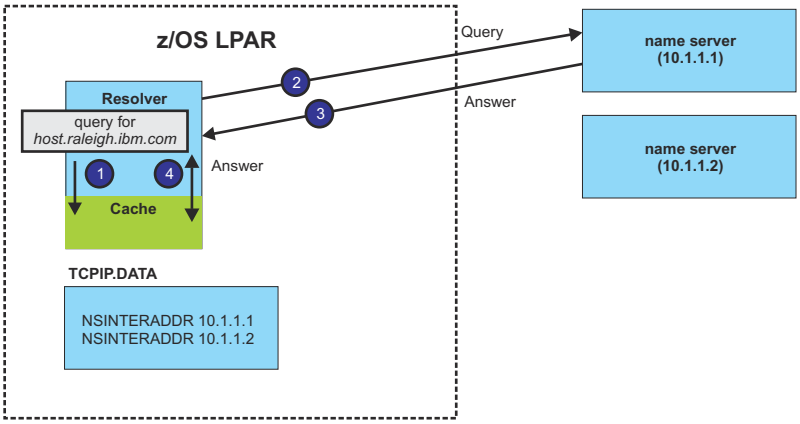


Figure 102. Resolver caching example

In the resolver caching model, when the initial query of the resolver cache (arrow 1) does not return any host name information, the DNS query is sent to the network name server (arrow 2), which might or might not already have information cached about host.raleigh.ibm.com. When the response is received (arrow 3), the information is cached by the resolver (arrow 4), and subsequent requests for host.raleigh.ibm.com can be satisfied without a DNS query (arrow 4).

Resolver caching is advantageous for the following reasons:

- You do not need to take any steps to enable resolver caching; it is automatically enabled. As shown in Figure 101 on page 750 and in Figure 102 on page 751, this eliminates the need to manually configure a local name server to cache the DNS responses.
- The cached information can be used by all applications that are running in the z/OS logical partition (LPAR), which provides the performance benefits of caching across the entire system for the cost of one DNS query.
- Resolver caching provides high performance because it reduces the network traffic to name servers.
- The resolver automatically regulates your storage use.
- Cached IP addresses can be returned and reordered in a round-robin manner, which provides a balanced way of using cached resources.

If you have been using a caching-only name server and you want to use resolver caching, see “Migrating from a local caching-only name server to resolver caching” on page 761.

You can disable resolver caching for selected applications; for more information, see “Steps for disabling caching for selected applications” on page 758.

Information that is cached by the resolver

Table 35 on page 751 shows the application programming interfaces (APIs) that use resolver caching.

Table 35. APIs that use resolver caching

API	Usage
getaddrinfo()	Resolves host name to one or more IP addresses. Supports both IPv4 and IPv6 addresses.
gethostbyaddr()	Resolves an IP address to a host name. Supports only IPv4 addresses.
gethostbyname()	Resolves a host name to one or more IP addresses. Supports only IPv4 addresses.

Table 35. APIs that use resolver caching (continued)	
API	**Usage**
getnameinfo()	Resolves an IP address to a host name. Supports both IPv4 and IPv6 addresses.

The resolver caches the following DNS response information generated by the APIs in Table 35 on page 751:

- Forward lookup information (IP addresses as A or AAAA records)

  Forward lookups are host-name-to-IP-address resolution requests. This includes IPv4 (A records) and IPv6 (AAAA records) addressing records from getaddrinfo() and gethostbyname() API calls.
- Reverse lookup information (domain name pointers as PTR records)

  Reverse lookups are IP-address-to-host-name resolution requests. This includes records from getnameinfo() and gethostbyaddr() API calls.
- Negative caching (NX) information

  Negative caching is the storage of the knowledge that a record does not exist, or that a request for a specific resource cannot or does not give an answer. The negative cache represents the received NXDOMAIN (nonexistent domain) responses from the server. The negative cache also represents the received NOERROR responses that did not include answer records of the requested type, such as a request for IPv6 addresses when the resource has only IPv4 addresses defined. Negative cache entries represent resources that are known to not exist and they include both reverse and forward entries.

Notes:

- IP addresses cached by getaddrinfo() and gethostbyname() API calls can be returned in a round-robin manner if CACHEREORDER is enabled.
- Sorting algorithms, such as the SORTLIST statement, might affect the results of the cache reordering logic. See "Interactions of address sorting and cache reordering" on page 756 for more information.

There is an upper limit on the amount of storage that can be used for negative caching information. The upper limit is 20 percent of the maximum amount of cache storage that the resolver is permitted to use. The resolver does not set aside this amount of storage for exclusive use by negative cache entries; rather, the resolver never exceeds this amount of storage to hold negative cache entries. When the upper limit is reached, no subsequent negative cache entries are saved until some existing entries are deleted and the negative cache entry storage use drops below the 20 percent upper limit. For information about setting the maximum amount of cache storage available to the resolver, see "Steps for configuring resolver caching (optional)" on page 757.

The resolver does not cache information retrieved from local host files, such as /etc/hosts and /etc/ipnodes; that type of information is already cached at a process level. The resolver also does not cache information retrieved using an API not listed in Table 35 on page 751.

The length of time that cache entries, including negative entries, are valid depends on the time-to-live (TTL) value that is returned by each domain name server. This is consistent with the behavior when using a caching-only name server or an intermediate name server to resolve a query.

In the following situations, cached entries are not saved up to the TTL value that is returned by the name server:

- When you flush the cache using the MODIFY RESOLVER,FLUSH,ALL command to delete all cached entries (for more information, see "Step for deleting cache entries" on page 760)
- When the maximum allocated storage for the cache is exhausted (for more information, see "Managing the cache size and cache storage" on page 759)
- When you limit the duration of time that any individual cache record can be saved using the MAXTTL resolver setup statement (for more information about the MAXTTL statement, see z/OS Communications Server: IP Configuration Reference)

- When you limit the duration of time that any individual negative cache record can be saved using the MAXNEGTTL resolver setup statement. For more information about the MAXNEGTTL statement, see z/OS Communications Server: IP Configuration Reference).

The organization of the cached data

The cache is contained in one physical location that has separate logical structures for forward and reverse lookup information. The cache data is organized by DNS name server, which permits different name servers to provide different values for a given host name or IP address.

For example, consider the installation shown in Figure 103 on page 753. There are two TCP stacks; one TCP stack is used for the core production processing, and a second TCP stack is used for test purposes only. The stacks use different name servers to isolate test resources from the production environment. Each name server can potentially have different IPv4 (A record) definitions for the same host name, as is the case for host.ibm.com. If the test application issues a query for host.ibm.com, the test TCP stack directs the request to the test name server, and IP address 10.45.5.5 is obtained. If the production application issues the same query, a different IP address (10.145.5.5) is obtained. The resolver caches both responses but remembers which response was received from which name server, so that a subsequent request from the test application for host.ibm.com returns the correct test IP address, and not the production IP address.

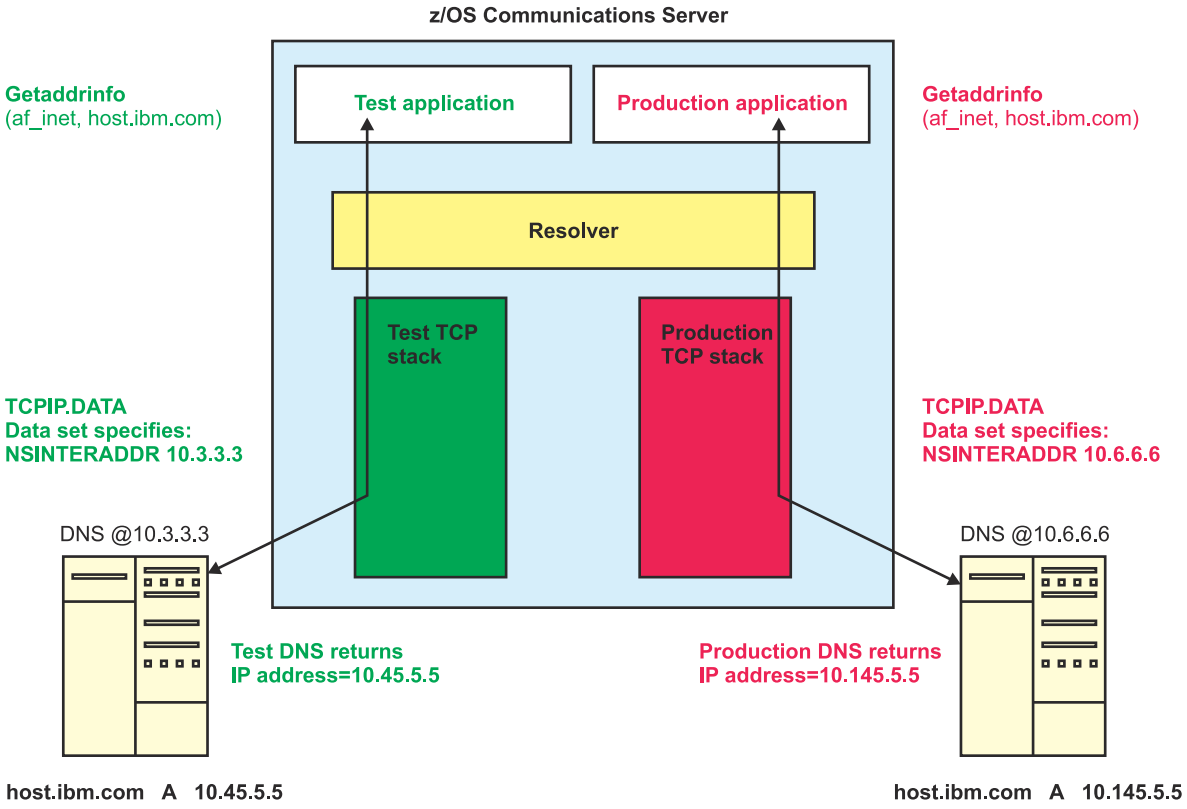


Figure 103. Resolver caching process; each stack specifies one NSINTERRADDR value

The resolver performs cache lookups using the NSINTERADDR search order list in the TCPIP.DATA data set. In Figure 103 on page 753, each stack specifies a single NSINTERADDR value. More likely, in reality, multiple NSINTERADDR definitions are provided, as shown in Figure 104 on page 754. In this example installation, if the primary name server is unavailable, a secondary name server is used instead. The resolver considers a cache entry that is associated with any name server in the NSINTERADDR list as a match for the target host name; the name resolution is complete and no DNS queries are sent to any of the name servers. The resolver searches the cache in the order that the name servers appear in the NSINTERADDR list; if multiple entries exist for the same target host name (one from each of the name servers in the NSINTERADDR list), the information provided by the first name server in the list is used. In Figure 104 on page 754, if the test application issues a query for host.ibm.com, IP address

10.145.5.5 is obtained because the first name server IP address in the list is now the production name server (10.6.6.6). If the production application issues the same query, the same IP address, 10.145.5.5 is obtained because there is no entry in the cache for host.ibm.com from the first name server in the list, but there is cache information from the second name server (which is, again, 10.6.6.6).

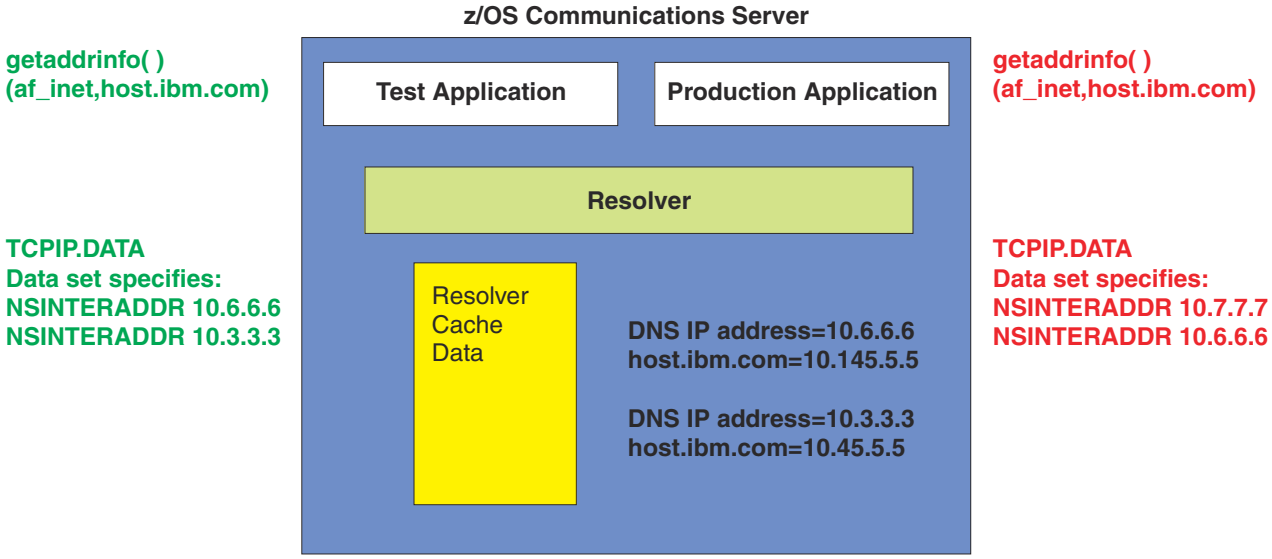


Figure 104. Resolver caching process; each stack specifies multiple NSINTERRADDR values

IPv4 information and IPv6 information are cached as separate entries. A maximum of 35 IP addresses is saved per host name for each name server that provides data for that host name.

Cache reordering

DNS servers can be configured to use a round-robin manner for ordering the IP addresses that are returned in response to forward lookup requests. The load balancing that this reordering provides is lost if the resolver does not also reorder its cached list of IP addresses in a round-robin manner. If the resolver does not perform cache reordering, the list of IP addresses is always returned in the same order that the information was received from the DNS server, as long as the cache information remains valid.

Starting in z/OS Communications Server V2R2, you can control whether the resolver reorders the list of cached IP addresses before the resolver returns the list in response to forward lookup resolution requests.

- Use the CACHEREORDER resolver setup statement to enable system-wide cache reordering.
- Use the NOCACHEREORDER resolver setup statement to disable system-wide cache reordering.

You can also disable cache reordering for an individual application by specifying NOCACHEREORDER in the TCPIP.DATA file that the application uses.

Rules:

- If both IPv6 and IPv4 addresses are cached for a host name, the two lists of addresses are managed and reordered independently of each other.
- The list of IP addresses is reordered independently of which applications issue the forward lookup requests.
- The list of IP addresses is reordered independently of the type of forward request that the application issues. The forward request can be getaddrinfo() or gethostbyname().
- If system-wide cache reordering is enabled, forward lookup requests from applications that have locally disabled cache reordering have no effect on the round-robin processing.
- If you dynamically change either system-wide or local setting of cache reordering to NOCACHEREORDER, the resolver, on subsequent forward lookup requests, returns the list of IP addresses in the order that the list was cached.

- The resolver performs the cache reordering function before any address sorting is performed. See "Interactions of address sorting and cache reordering" on page 756 for more information.
- Cache reordering has no effect on reverse lookup requests.

Example

Consider host name **reorder.domain** that has three IPv6 and four IPv4 addresses that were cached in the following order:

```
    IPv6 addresses   2001::10:9:1:1
                     2001::10:9:1:2
                     2001::10:9:1:3
    IPv4 addresses   10.9.1.1
                     10.9.1.2
                     10.9.1.3
                     10.9.1.4
```

If system-wide cache reordering is in effect, the resolver reorders the lists as follows for this sequence of forward lookup requests for **reorder.domain**:

1. Application APPLA issues a getaddrinfo() request for IPv4 addresses:

```
    IPv4 addresses   10.9.1.2
                     10.9.1.3
                     10.9.1.4
                     10.9.1.1
```

2. Application APPLD issues a getaddrinfo() request for both IPv4 and IPv6 addresses:

```
    IPv6 addresses   2001::10:9:1:2
                     2001::10:9:1:3
                     2001::10:9:1:1
    IPv4 addresses   10.9.1.3
                     10.9.1.4
                     10.9.1.1
                     10.9.1.2
```

3. Application APPLNOCR, which has NOCACHEREORDER specified in its TCPIP.DATA file, issues a getaddrinfo() request for both IPv4 and IPv6 addresses:

```
    IPv6 addresses   2001::10:9:1:1
                     2001::10:9:1:2
                     2001::10:9:1:3
    IPv4 addresses   10.9.1.1
                     10.9.1.2
                     10.9.1.3
                     10.9.1.4
```

4. Application APPLB issues a gethostbyname() request for IPv4 addresses:

```
    IPv4 addresses   10.9.1.4
                     10.9.1.1
                     10.9.1.2
                     10.9.1.3
```

5. Application APPLC issues a getaddrinfo() request for IPv6 addresses:

```
    IPv6 addresses   2001::10:9:1:3
                     2001::10:9:1:1
                     2001::10:9:1:2
```

6. Application APPLB issues a getaddrinfo() request for both IPv4 and IPv6 addresses:

```
    IPv6 addresses   2001::10:9:1:1
                     2001::10:9:1:2
                     2001::10:9:1:3
    IPv4 addresses   10.9.1.1
                     10.9.1.2
                     10.9.1.3
                     10.9.1.4
```

7. Application APPLB issues a gethostbyname() request for IPv4 addresses:

```
    IPv4 addresses   10.9.1.2
                     10.9.1.3
                     10.9.1.4
                     10.9.1.1
```

8. Application APPLD issues a getaddrinfo() request for both IPv4 and IPv6 addresses:

```
    IPv6 addresses   2001::10:9:1:2
                     2001::10:9:1:3
                     2001::10:9:1:1
    IPv4 addresses   10.9.1.3
                     10.9.1.4
                     10.9.1.1
                     10.9.1.2
```

Notes:

- Because cache reordering does not consider which applications issue the forward lookup requests, an individual application might get results that suggest that cache reordering is not being performed even though it is being performed. For example, the two requests from APPLD result in the same lists even though requests from other applications result in lists with different orders.
- Because NOCACHEREORDER is specified in its TCPIP.DATA file, the resolver returns the list of IP addresses in the same order for every request APPLNOCR issues. The requests from APPLNOCR have no effect on the other results where cache reorder is performed.

Interactions of address sorting and cache reordering

The resolver can perform the following types of IP address sorting after the resolver obtains the list of IP addresses for a forward lookup request:

- IPv4 addresses are sorted based on the values that are defined on the SORTLIST statement in the application's TCPIP.DATA file. To prevent this kind of sorting, do not specify a SORTLIST statement.
- As part of getaddrinfo() processing, IPv4 and IPv6 addresses are sorted based on the default address selection algorithm. You cannot prevent this sorting from being performed. However, to influence the algorithm, you can define a policy table by using the DEFADDRTABLE statement in the TCP/IP profile. See DEFADDRTABLE statement in z/OS Communications Server: IP Configuration Reference for more information.

When both cache reordering and address sorting are performed, the resolver first reorders the list of IP addresses. The appropriate sorting algorithm is then performed on the reordered list of IP addresses. Depending on the sorting algorithm and the IP addresses in the list, the outcome of the sorting algorithm might cause the resolver to return the list of IP addresses in the same order for every forward lookup request.

For example, if host name **sorted.domain** resolves to IPv4 addresses 10.9.1.1 and 20.10.8.7, the following results might occur:

- If no SORTLIST statement is coded, half of the resolution requests for **sorted.domain** have IP address 10.9.1.1 listed first, and half of the requests have IP address 20.10.8.7 listed first.
- If a SORTLIST statement is coded, the results depend on how the sorting algorithm applies to these addresses:
  - If the sorting values on the SORTLIST statement favor one address over the other, the chosen address is always displayed first in the returned list, regardless of cache reordering. For example, if 10.9.1/24 is coded on the SORTLIST statement, IP address 10.9.1.1 is favored and is always first in the sorted list.
  - If the sorting values on the SORTLIST statement do not favor any address, the results are similar to when no SORTLIST statement is defined. For example, if 55.1/16 is coded on the SORTLIST statement, no IP address is favored. Half of the requests have IP address 10.9.1.1 listed first and half have IP address 20.10.8.7 listed first.

Steps for configuring resolver caching (optional)

Resolver caching is automatically enabled and you do not need to make parmlib or JCL changes. Unless you choose to do so, you do not have to configure the cache; however, you can explicitly configure resolver caching.

Before you begin

You must have already created a resolver setup file; see "The resolver setup file" on page 738 for more information.

Procedure

Perform the following steps to configure resolver caching:

1. Specify the CACHE statement to enable system-wide caching of Domain Name System (DNS) queries that have been resolved.

   System-wide caching is enabled by default, but you can explicitly enable it by specifying this statement.
2. Specify the CACHESIZE(*cachesize*M) statement to define the amount of storage, in megabytes, that can be allocated by the resolver to manage cached records.

   The default value is 200 megabytes.

   Guideline: If you set a CACHESIZE value that is too low, the resolver might repeatedly take action to reduce cache usage. You should set the CACHESIZE value to be at least 50 percent higher than your expected usage.

   If you specify this statement and the NOCACHE statement, the CACHESIZE statement is ignored. For more information about cache size, see "Managing the cache size and cache storage" on page 759.
3. Specify the MAXTTL statement to define the maximum amount of time, in seconds, that cache entries are considered to be valid by the resolver.

   The default value is the maximum time-to-live (TTL) value that is provided by the DNS name server for this resource. If you specify this statement and the NOCACHE statement, the MAXTTL statement is ignored.
4. Specify the MAXNEGTTL statement to define the maximum amount of time, in seconds, that negative cache entries are considered to be valid by the resolver. The default value is the setting for MAXTTL if it is specified, otherwise the default is the maximum TTL value that is provided by the DNS name server for the resource. If you specify this statement and the NOCACHE statement, the MAXNEGTTL statement is ignored.
5. Specify whether cache reordering is used on a system-wide basis.
   - Specify CACHEREORDER to enable system-wide cache reordering.
   - Specify NOCACHEREORDER to disable system-wide cache reordering.

   The default setting is to disable system-wide cache reordering. If you specify the NOCACHE statement, the CACHEREORDER and NOCACHEREORDER statements are ignored. See "Cache reordering" on page 754 for more information.
6. Perform one of the following steps:
   - If the resolver is not active, start the resolver.
   - If the resolver is currently active, issue the MODIFY RESOLVER,REFRESH,SETUP=*setup_file_name* command to cause the resolver to use the new settings.

Results

You complete the task when the correct values for the CACHE, CACHEREORDER or NOCACHEREORDER, CACHESIZE(*cachesize*M), MAXTTL and MAXNEGTTL statements are displayed after you issue the START command or in the MODIFY RESOLVER,REFRESH command output.

For more information about resolver setup statements, see z/OS Communications Server: IP Configuration Reference.

For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.

Steps for disabling caching for selected applications

You can disable resolver caching for applications in your environment that do not need to use it.

Before you begin

Some applications might not need to use resolver caching. For example, as shown in Figure 103 on page 753, you might have a production network and a test network in your environment. Users of the test network might require specialized host name resolution using a unique set of DNS servers that are not used in the production network. Because the test network is isolated and is likely to affect only a small number of users, using resolver caching for the test network could waste cache storage, and could complicate operation of the production network by adding information that is not pertinent to the production network. You can disable resolver caching for applications using the test network, while continuing to use resolver caching for the production network.

Procedure

Perform the following steps to disable caching for some applications:

1. Identify or create the TCPIP.DATA data set associated with the application for which you want to disable resolver caching.
2. Turn off the resolver caching function by specifying the NOCACHE statement in that TCPIP.DATA data set.
3. Issue the MODIFY RESOLVER,REFRESH command to cause the resolver to refresh the settings for the application.
4. Activate the Trace Resolver facility to determine which TCPIP.DATA values are being used by the resolver and where they are being read from.
5. When the trace is active, issue the Netstat HOME/-h command to display the values.

Results

The task is completed when the value NOCACHE is displayed in the Trace Resolver output that is generated by the Netstat HOME/-h command.

For more information about configuration statements in TCPIP.DATA, see z/OS Communications Server: IP Configuration Reference. For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.

Steps for disabling cache reordering for selected applications

You can disable cache reordering for applications in your environment if you do not need to use it.

Before you begin

Some applications might not need cache reordering. For example, applications that do not have load balancing requirements or that have that have a well-defined algorithm might require the resolver to return the cached addresses in the same order all the time. Determine which applications do not require cache reordering and disable cache reordering for these specialized applications.

Procedure

Perform the following steps to disable cache reordering for the selected applications:

1. Identify or create the TCPIP.DATA data set that is associated with the application for which you want to disable cache reordering.

2. Turn off the cache reordering function by specifying the NOCACHEREORDER statement in the TCPIP.DATA data set.
3. Issue the MODIFY RESOLVER,REFRESH command to refresh the settings for the application.
4. Activate the trace resolver facility to determine the TCPIP.DATA values that the resolver is using and where the values are read from.
5. When the trace resolver is active, issue the Netstat HOME/-h command to display the values.

Results

Cache reordering is disabled when the value NOCACHEREORDER is displayed in the trace resolver output that the Netstat HOME/-h command generates.

For more information about configuration statements in the TCPIP.DATA data set, see z/OS Communications Server: IP Configuration Reference. For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.

Managing the cache size and cache storage

Resolver cache information is maintained in 64-bit private storage in the resolver address space, which is also referred to as the storage above the 2 GB bar. This means that caching does not impact common storage.

The default maximum size of the cache is 200 megabytes. The size is allocated incrementally, as the cache increases. For planning purposes, assume that 1 megabyte of storage holds between 400 and 450 cache entries. The actual number of cache entries depends on the amount of storage used for cache infrastructure control blocks, which varies depending on the number of name servers and the number of entries. If you want to control the maximum size of the cache instead of using the 200 megabyte default, you can use the CACHESIZE resolver setup statement. The CACHESIZE statement specifies the maximum amount of storage, in megabytes, that can be allocated by the resolver to manage cached records. The CACHESIZE statement is also used to determine the upper limit of storage that can be used for negative cache entries. If you use the default CACHESIZE value of 200M, then no more than 40 megabytes of that storage will be used for negative cache entries.

The resolver does not automatically delete expired records (those records whose TTL value has been exceeded). Regardless of the amount of cache storage that is in use, the resolver deletes expired records if a new request is received for the expired resource. The resolver also deletes expired records when your cache storage usage has reached the following levels, and takes additional actions to reduce your storage usage.

- When cache storage is less than 75 percent full, the resolver deletes all expired records approximately every 10 minutes, without waiting for a new request to be processed.
- When cache storage is 75 - 97 percent full, the resolver deletes all expired records approximately every minute, without waiting for a new request to be processed.
- When cache storage is 98 percent or more full, the resolver deletes all expired records approximately every 30 seconds, without waiting for a new request to be processed. The resolver does not add new records to the cache while usage is greater than 99 percent. Message EZZ9307E is displayed until usage is less than 90 percent, you increase the CACHESIZE value, or you delete the contents of the cache; see “Step for deleting cache entries” on page 760 for more information.

Steps for manually managing the storage capacity of the resolver cache

You can increase the size of the cache, decrease the length of time that cache records are saved in the cache, and delete all cache entries.

Before you begin

You must have already created a resolver setup file; see “The resolver setup file” on page 738 for more information.

Procedure

Perform one of the following steps to manage the storage capacity of the resolver cache:

- On the CACHESIZE statement in the resolver setup file, increase the CACHESIZE value and issue the MODIFY RESOLVER,REFRESH,SETUP=*ressetup_filename* command.
- On the MAXTTL statement in the resolver setup file, decrease the MAXTTL value to decrease the length of time that cache records are saved in the cache and issue the MODIFY RESOLVER,REFRESH,SETUP=*ressetup_filename* command. The new MAXTTL value affects only new cache records as they are created; there is no effect on existing cache records.
- On the MAXNEGTTL statement in the resolver setup file, or if you had previously been taking the value of MAXTTL for the default for MAXNEGTTL, decrease or specify a smaller MAXNEGTTL value to decrease the length of time that negative cache records are saved in the cache. Issue the MODIFY RESOLVER,REFRESH,SETUP=*ressetup_filename* command. The new MAXNEGTTL value affects only new negative cache records as they are created; there is no effect on existing cache records.

Note: You can completely disable caching of negative cache records by specifying a value of 0 for MAXNEGTTL.

- Issue the MODIFY RESOLVER,FLUSH,ALL command to delete all cache entries.

Results

You know you are done when message EZZ9307E is no longer displayed on the operator console.

Step for deleting cache entries

You might want to delete the contents of the cache when an IP address has changed for a given host name or when your cache storage is exhausted.

Procedure

Perform the following step to delete the contents of the cache:

- Issue the MODIFY RESOLVER,FLUSH,ALL command to delete all cache entries.

  For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.

Results

If you have deleted cache entries because your cache storage was exhausted, you know you are done when message EZZ9307E no longer appears on the screen.

Step for displaying the contents of the cache

Procedure

Perform the following step to display the contents of the resolver cache:

- Issue the Netstat RESCache/-q command to display the RESCache/-q report.

  The Netstat RESCache/-q report displays the contents of the resolver cache data. You can display statistical data, both overall and by name server. You can also create a report that shows all the cache entries, or you can use filters to produce a report that shows subsets of cache entries.

What to do next

For details about access control considerations for the Netstat RESCache/-q report, see “Netstat access control” on page 145. For Netstat command syntax and sample report output, see z/OS Communications Server: IP System Administrator's Commands.

Migrating from a local caching-only name server to resolver caching

If you have been using a local caching-only name server, you should consider using resolver caching. You can compare the contents of your local caching-only name server with the contents of the resolver cache by performing these steps.

Before you begin

Resolver caching must be enabled and active. If resolver caching is not enabled (the NOCACHE statement is coded in the resolver setup statement), see "Steps for configuring resolver caching (optional)" on page 757.

Procedure

Perform the following steps to migrate from a local caching-only name server to resolver caching:

1. Display the contents of your caching-only name server and of the resolver cache at specific intervals.
   - For your caching-only name server, you can dump the contents of the DNS cache. For DNS servers running BIND 9, use the **rndc dumpdb** command.
   - For the resolver cache, use the Netstat RESCache/-q report. For more information, see "Step for displaying the contents of the cache" on page 760.
2. Compare the contents of the caching-only name server and the resolver cache, and determine whether to use only resolver caching or resolver caching with the local caching-only name server, using the following criteria:
   - If the contents are similar, and consist primarily of A, AAAA, and PTR DNS records, then you would benefit the most by using only resolver caching. This situation is the most common.
   - If the contents are dissimilar, but the caching-only name server has primarily A, AAAA, and PTR DNS records, then the dissimilar contents are most likely the result of differences in how the resolver cache and the caching-only name server delete expired records. In this situation, you are still most likely to benefit from using only resolver caching.
   - If the contents are dissimilar, and the caching-only name server has many DNS records that are not A, AAAA, or PTR records, you will probably benefit the most by using both resolver caching and the caching-only name server. This situation is not common.
3. Calculate the amount of resolver cache storage that you think you need.

   You can use the default amount of storage (200 megabytes) or you can use the CACHESIZE resolver setup statement to specify a maximum amount of storage. For calculation purposes, 1 megabyte of storage holds roughly 400 - 450 cache entries.
4. If you are not going to use the local caching-only name server, stop that server.

   Guideline: If the local caching-only name server is the only name server in the NSINTERADDR list of name servers to be contacted, replace the caching-only name server entry with one or more name server IP addresses to be contacted. If there is already more than one name server in the NSINTERADDR list of name servers, delete the IP address of the local caching-only name server.

Results

For more information about resolver setup statements, see z/OS Communications Server: IP Configuration Reference.

Monitoring the responsiveness of Domain Name System name servers

The resolver monitors the responsiveness level of Domain Name System (DNS) name servers that are in the network. The resolver provides two levels of name server monitoring:

- "Network operator notification" on page 763

  The resolver alerts the network operator about name servers that fail to respond to a significant percentage of resolver queries, but continues to send DNS queries that are generated by an application

to the unresponsive name server. You can use these alerts to better manage the list of name servers that the system uses and to avoid unnecessary delays when a host name or IP address is being resolved.

- "Autonomic quiescing of unresponsive name servers" on page 764

  The resolver alerts the network operator about name servers that fail to respond to a significant percentage of resolver queries, and does not send additional DNS queries that are generated by an application to the unresponsive name server. While the name server remains unresponsive, the resolver periodically sends DNS polling queries to the name server. When the name server is responsive to the resolver's DNS polling queries, the resolver resumes sending DNS queries that are generated by an application to the name server.

To determine name server responsiveness, the resolver collects statistics about name server responsiveness at 30-second or 1-minute intervals, depending on the monitoring function that is being performed. During a given monitoring interval, the resolver keeps system-wide statistics about the total number of resolver queries that are sent to a name server and about the number of those resolver queries that were not responded to by the name server. At the end of the monitoring interval, the resolver calculates a percentage of the total number of queries that were not responded to by the name server over the course of the last interval, or the last five intervals, depending on the monitoring function being performed. This percentage is compared to the setting on the UNRESPONSIVETHRESHOLD resolver setup statement. If the percentage of failures equals or exceeds the threshold value, the resolver considers the name server to be unresponsive. For information about the UNRESPONSIVETHRESHOLD statement and how to set its value, see z/OS Communications Server: IP Configuration Reference and "Optimizing the UNRESPONSIVETHRESHOLD value for your network" on page 768.

The phrase *resolver queries* does not mean the same thing as *resolver API calls* in the context of name server responsiveness. A single resolver API call, such as getaddrinfo() or gethostbyname(), can generate multiple resolver queries to one or more DNS name servers, based on retry counts, domain names to append to a search, or the type of information that is being requested by the API. Conversely, a resolver API call might not generate any resolver queries to any DNS name servers, if the resource is already in the resolver cache. See "Examples of resolver monitoring of DNS name servers" on page 766 for examples of how different TCPIP.DATA file settings can influence name server responsiveness statistics.

Restriction: The resolver can monitor a maximum of 32 name servers for responsiveness.

The resolver considers the following failures to be indicative of an unresponsive name server:

- The resolver sends a UDP or TCP query to a name server and never receives a response.
- The resolver sends a UDP query to a name server and receives a response after the RESOLVERTIMEOUT value has expired.
- The resolver attempts to send data to a name server using UDP, but the data cannot be sent to the target IP address (for example, because of an error in the route configuration).
- The resolver attempts to connect to a name server using TCP, but the connection attempt times out.
- In some situations, the BIND 9 DNS utilities (for example, dig or nsupdate) issue getaddrinfo() API calls to resolve a host name that represents a remote DNS name server, and those API calls invoke z/OS resolver processing. If any of the previously mentioned failures occur during these BIND 9 resolver calls, the failures are included in the name-server statistics.

The resolver does not consider the following failures to be indicative of an unresponsive name server:

- The resolver cannot open a socket (UDP or TCP) to send a request to a name server, including instances in which the system is IPv4-only capable and an IPv6 name server IP address is coded on the NSINTERADDR statement.
- The resolver sends a UDP query to a name server to determine whether the name server is EDNS0-capable, but does not receive a response to that UDP query; see "Extension Mechanisms for DNS standards and the resolver" on page 770 for more information about EDNS0 processing.
- The resolver sends a UDP query to a name server and the name server responds with a DNS return code (such as SERVFAIL or NOTIMPL) that indicates that the name server is active and responding but is unable to process the request that was sent.

- Timeouts or failures occur during BIND 9 DNS utility processing that does not involve getaddrinfo() calls (that processing uses BIND 9 resolver services to send queries to a name server).

Network operator notification

You can specify only a failure threshold percentage using the UNRESPONSIVETHRESHOLD resolver setup statement, or you can use the default value, which is 25%. In either case, the resolver reacts to unresponsive name servers by notifying the network operator of the condition and continues to send DNS queries to the unresponsive name server. The network operator notification function collects name-server statistics at 1-minute intervals, but makes decisions based on the last five monitoring intervals. These intervals are called sliding 5-minute windows.

Messages generated by the resolver for the network operator notification function

If the resolver detects that a name server is not being responsive, a series of network operator messages is issued that relate to that name server. For example, if a name server is operating at IP address 10.42.35.200 and the UNRESPONSIVETHRESHOLD value is 25, then the following sequence of messages might be generated by the resolver:

1. At the end of a 5-minute monitoring interval, the resolver determines that the name server failed to respond to 35% of 6000 queries that were attempted by the resolver. The resolver considers the name server to be unresponsive and issues the following messages:

   ```
   EZZ9308E UNRESPONSIVE NAME SERVER DETECTED AT IP ADDRESS 10.42.35.200
   EZZ9310I NAME SERVER 10.42.35.200
        TOTAL NUMBER OF QUERIES SENT    6000
        TOTAL NUMBER OF FAILURES        2100
        PERCENTAGE                        35%
   ```

2. At the end of the next 5-minute interval, the resolver determines that the name server failed to respond to 55% of the 3000 queries that it attempted during that interval. The name server is still considered to be unresponsive, and the following message is issued:

   ```
   EZZ9310I NAME SERVER 10.42.35.200
        TOTAL NUMBER OF QUERIES SENT    3000
        TOTAL NUMBER OF FAILURES        1650
        PERCENTAGE                        55%
   ```

   This message and the statistical information for the name server are issued at 5-minute intervals for as long as the resolver considers the name server to be unresponsive.

3. At the end of a subsequent monitoring interval, the resolver determines that the name server failed to respond to 15% of the 4500 queries that the resolver attempted during the latest sliding 5-minute interval. This percentage is under the threshold value, so the resolver considers this name server to be responsive again. The resolver clears message EZZ9308E from the operator console and issues the following messages:

   ```
   EZZ9309I NAME SERVER IS NOW RESPONSIVE AT IP ADDRESS 10.42.35.200
   EZZ9310I NAME SERVER 10.42.35.200
        TOTAL NUMBER OF QUERIES SENT    4500
        TOTAL NUMBER OF FAILURES         675
        PERCENTAGE                        15%
   ```

4. The resolver also clears message EZZ9308E from the operator console if any of the following situations occur:

   - The network operator disables the monitoring function using the MODIFY RESOLVER,REFRESH,SETUP command. For more information, see “Steps for modifying the UNRESPONSIVETHRESHOLD value” on page 769.
   - The network operator enables the autonomic quiescing of unresponsive name servers function using the MODIFY RESOLVER,REFRESH,SETUP command. For more information, see “Steps for modifying the UNRESPONSIVETHRESHOLD value” on page 769.
   - No resolver queries are sent to the name server during the current 5-minute monitoring interval.
   - The network operator notification function abnormally terminates due to system error.

- The resolver is stopped.

The resolver issues individual messages for each name server that it considers to be unresponsive at the end of a given monitoring interval. Because the resolver calculates the responsiveness of a name server by using a sliding 5-minute interval, messages might be issued for different name servers at different times, rather than notifications about all unresponsive name servers being issued at the same time.

Diagnosing problems with unresponsive name servers

Use the statistics supplied in message EZZ9310I to determine the severity of the problem that the resolver is reporting. For example, if the unresponsive name server is the primary server for the network, and it is failing to respond to over half of a large number of queries that are directed to it, that situation might be more serious than a secondary name server that is not responding to 25% of a small number of queries. You can also use the statistics that are displayed when message EZZ9309I is issued to determine whether the name server is fully responsive again, is likely to become unresponsive again because the failure rate is still close to the threshold value, or is just not being used in your environment.

Guideline: The resolver does not require that there be a certain number of failed queries to a particular name server before it declares that server to be unresponsive. For example, if only one query is received by a particular name server during a monitoring interval, and that query fails to obtain a response, then the resolver considers that name server to be 100% unresponsive.

Autonomic quiescing of unresponsive name servers

In addition to specifying a failure threshold percentage on the UNRESPONSIVETHRESHOLD resolver setup statement, you can also specify the AUTOQUIESCE operand. If you specify the AUTOQUIESCE operand, in addition to network operator notification, the resolver stops sending DNS queries that are generated by an application to the unresponsive name server. While a name server is considered to be unresponsive, the resolver sends DNS polling queries to the name server every 6 seconds. When the name server is responsive to the resolver's DNS polling queries, the resolver resumes sending DNS queries that are generated by an application to the name server. The autonomic quiescing of unresponsive name servers function collects name-server statistics in intervals of 30 seconds, but makes decisions based on no more than two monitoring intervals.

Guideline: If all name servers specified on the TCPIP.DATA NSINTERADDR statements are unresponsive, the resolver sends DNS queries that are generated by an application to those name servers; the resolver does not fail the application queries immediately.

Restriction: If you specify the AUTOQUIESCE operand on the UNRESPONSIVETHRESHOLD statement, you must also specify the GLOBALTCPIPDATA statement. If you do not specify the GLOBALTCPIPDATA statement, the resolver issues message EZD2036I and ignores the AUTOQUIESCE operand. The resolver operates as if you requested the network operator notification function. For more information about the UNRESPONSIVETHRESHOLD statement, see z/OS Communications Server: IP Configuration Reference.

If you use multiple TCP/IP stacks in a common INET (CINET) environment, you must ensure that the IP addresses of the name servers that are specified in the global TCPIP.DATA file are accessible from all TCP/IP stacks. If a name server IP address is accessible only from a specific TCP/IP stack, the autonomic quiescing function could mistakenly consider that name server as unresponsive to DNS queries. For example, consider two TCP/IP stacks, TCPIP1 and TCPIP2, where the IP address for DNS1 is accessible using only TCPIP1 and the IP address for DNS2 is accessible using only TCPIP2. If the NSINTERADDR statement in the global TCPIP.DATA specifies DNS1 and DNS2, in that order, any application DNS queries with stack affinity to TCPIP2 results in DNS query failures to DNS1 because the IP address for DNS1 is not accessible using TCPIP2. If there is a sufficient volume of these queries with stack affinity to TCPIP2, or the UNRESPONSIVETHRESHOLD value is very small, the resolver might decide that DNS1 is unresponsive and stop using it temporarily, resulting in unnecessary failures for application DNS queries that use TCPIP1.

Guideline: If you cannot ensure that all DNS IP addresses are accessible from all your TCP/IP stacks, you should not use a global TCPIP.DATA file, and you should not code AUTOQUIESCE on the UNRESPONSIVETHRESHOLD setup statement in your resolver setup file.

Messages generated by the resolver for the autonomic quiescing of unresponsive name servers function

If the resolver detects that a name server is not being responsive, a series of network operator messages is issued that relate to that name server. For example, if a name server is operating at IP address 10.42.35.200 and the UNRESPONSIVETHRESHOLD value is 100, then the following sequence of messages might be generated by the resolver:

1. At the end of a 30-second monitoring interval, the resolver determines that the name server failed to respond to 100% of the 500 queries that were attempted by the resolver. The resolver considers the name server to be unresponsive and issues the following messages:

   ```
   EZZ9311E STOPPED USING NAME SERVER AT IP ADDRESS 10.42.35.200
   EZZ9313I NAME SERVER 10.42.35.200
        TOTAL NUMBER OF QUERIES SENT         500
        TOTAL NUMBER OF FAILURES             500
        TOTAL NUMBER OF RESOLVER POLLS SENT    0
        TOTAL NUMBER OF POLL FAILURES          0
        PERCENTAGE                           100%
   ```

   After issuing this message, the resolver sends only DNS polling queries to the unresponsive name server.
2. At the end of a subsequent monitoring interval, the resolver determines that the name server failed to respond to 10% of the 10 or more DNS polling queries during the last two monitoring intervals. This percentage is below the 100% threshold value, so the resolver considers this name server to be responsive again. The resolver clears message EZZ9311E from the operator console and issues the following message:

   ```
   EZZ9312I RESUMED USING NAME SERVER AT IP ADDRESS 10.42.35.200
   ```
3. After the resolver issues message EZZ9312I, it resumes sending DNS queries that are generated by an application to the name server.
4. The resolver also clears message EZZ9311E from the operator console if any of the following events occur:
   - The network operator disables the monitoring function using the MODIFY RESOLVER,REFRESH,SETUP command. For more information, see “Steps for modifying the UNRESPONSIVETHRESHOLD value” on page 769.
   - The network operator enables the network operator notification function using the MODIFY RESOLVER,REFRESH,SETUP command. For more information, see “Steps for modifying the UNRESPONSIVETHRESHOLD value” on page 769.
   - The autonomic quiescing of unresponsive name servers function abnormally terminates because of a system error.
   - The resolver is stopped.

The resolver issues individual messages for each name server that it considers to be unresponsive at the end of a given monitoring interval. Because the resolver calculates the responsiveness of a name server at every interval, messages might be issued for different name servers at different times, rather than notifications being issued about all unresponsive name servers at the same time. The resolver does not periodically reissue message EZZ9313I for each name server, but the status and current failure rate for each name server specified on an NSINTERADDR statement in the global TCPIP.DATA file is included in the MODIFY RESOLVER,DISPLAY command output when the autonomic quiescing function is active. The failure rate that is displayed is the percentage calculated by the resolver for the individual name server at the end of the last 30-second monitoring interval. For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.

Guideline: In most cases, the resolver requires a sample size of at least 10 DNS queries before it declares a name server to be either unresponsive or responsive. For example, if only one query is received by a particular name server during a 30-second monitoring interval, and that query fails to obtain a response, then the resolver does not consider that name server to be unresponsive, even though the failure rate is 100% and exceeds the UNRESPONSIVETHRESHOLD value. The resolver sends DNS polling queries to

this name server to collect a sufficient sample size of responses. In subsequent intervals, the resolver uses the results from DNS queries that are generated by an application and from DNS polling queries to calculate the overall failure rate for the name server.

How the resolver polls unresponsive name servers

The resolver sends DNS polling queries to unresponsive name servers every 6 seconds. The resolver uses the value specified on the RESOLVERTIMEOUT statement in the global TCPIP.DATA file to determine how long to wait for a response to a given DNS polling query. If the RESOLVERTIMEOUT value is greater than 6, multiple DNS polling queries can be outstanding to a given name server.

Rule: If you change the value of the RESOLVERTIMEOUT statement, only the timeout value of subsequent DNS polling queries is affected. For more information about the RESOLVERTIMEOUT statement, see z/OS Communications Server: IP Configuration Reference.

The resolver continues to poll unresponsive name servers until one of the following events occur:

- The name server responds to all DNS queries generated by an application and to all DNS polling queries.

  If a name server is unresponsive, the only queries that the resolver sends to the name server are DNS polling queries. The resolver sends DNS queries generated by an application to the name server as soon as the name server is responsive to DNS polling queries, but the resolver continues polling a name server as long as some DNS queries result in failures. The resolver considers a name server to be responsive to DNS polling queries if one of the following situations occur:

  – If the name server had responded to DNS queries that are generated by an application prior to becoming 100% unresponsive to these queries, a single positive response to a DNS polling query is sufficient for the name server to be considered responsive.
  – In all other cases, the name server must respond to a sufficient percentage of DNS polling queries to ensure that the following conditions are true:
    - At least 10 DNS polling queries received positive responses or timed out within the last 60 seconds
    - The failure rate for this name server in terms of responding to DNS polling queries is less than the UNRESPONSIVETHRESHOLD percentage.

  Any response from the name server to a DNS polling query is considered to be a positive response, including responses that specify NXDOMAIN, SERVFAIL, or similar responses.
- The autonomic quiescing of unresponsive name servers is stopped with a MODIFY REFRESH,SETUP command. For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.
- The name server is removed from the list of name servers that is specified on the NSINTERADDR statements in the global TCPIP.DATA file. The resolver stops polling this name server whether the server was explicitly removed from the list by the network operator using a MODIFY REFRESH command or implicitly removed as a result of a DNS query generated by an application. For more information about the MODIFY command for the resolver address space, see z/OS Communications Server: IP System Administrator's Commands.
- The resolver is stopped.

Examples of resolver monitoring of DNS name servers

Values in the TCPIP.DATA file can affect the statistics that the resolver collects when it monitors DNS name servers. For example, consider the following settings from TCPIP.DATA:

```
NAMESERVER 10.43.25.200 10.43.125.203 10.43.25.200
RESOLVERUDPRETRIES 2
RESOLVERTIMEOUT 0.075
RESOLVEVIA UDP
```

In this example, one name server (10.43.25.200) appears twice in the list of name servers that the resolver will search. The resolver should try that list of name servers again one time before it considers

the name servers to be unresponsive. Assume that the resolver generates a query to resolve the address user.ibm.com as part of gethostbyname processing. The following example sequence occurs:

1. The resolver sends the query to name server 10.43.25.200, which times out after 75 milliseconds (based on the RESOLVERTIMEOUT value).
2. The resolver forwards the request to name server 10.43.125.203, which also times out.
3. The request goes to name server 10.43.25.200 a second time (as the last name server in the list), which times out again.

   The first retry of the list name servers is complete.
4. The resolver begins at the top of the list again and sends the request to name server 10.43.25.200 for a third time. A response arrives from this name server, possibly as the result of name server delays.
5. The resolver stops searching for the resource.

Based on the search that the resolver performed, the system-wide total request count for name server 10.43.25.200 is incremented by 3, and the total failure count is incremented by 2. If the searches that are shown in this example are all the activity for this name server over the course of a 5-minute sliding window or the 30-second monitoring interval, the failure rate for this name server is 66%; the system-wide total request count and the total failure count for name server 10.43.124.203 are both incremented by 1. If the resolver does not send any more queries to this name server during the 5-minute sliding window or the 30-second monitoring interval, the failure rate for name server 10.43.124.203 is 100%.

If you are using the network operator notification and the threshold percentage is less than 66%, the resolver reports both of these name servers as unresponsive, but continues to send DNS queries to both name servers.

If you are automatically quiescing unresponsive name servers, the resolver does not consider either of these name servers to be unresponsive because less than 10 queries were directed to the name server during the 30-second interval. The name servers are not considered unresponsive regardless of the setting of the UNRESPONSIVETHRESHOLD value.

Consider these different TCPIP.DATA file settings:

```
NAMESERVER 10.43.25.200
SEARCH raleigh.ibm.com
RESOLVERTIMEOUT 0.075
RESOLVEVIA UDP
```

In this example, only one name server is coded, and only one domain name can be appended to the input host name as an additional search attempt. Assume that an application issues getaddrinfo() for host name `user`, and that ai_family=AF_UNSPEC is specified. The following example sequence occurs:

1. The resolver searches for domain name `user.raleigh.ibm.com` and requests AAAA records.
2. One of the following actions occurs:
   - If the resolver obtains resource information, the search ends.
   - If the resolver does not obtain resource information, the resolver continues to request AAAA records, but searches the next domain in the sequence, which is `user`.
3. One of the following actions occurs:
   - If the resolver obtains resource information, the search ends.
   - If the resolver does not obtain resource information, the resolver searches for domain name `user.raleigh.ibm.com` and requests A records.
4. One of the following actions occurs:
   - If the resolver obtains resource information, the search ends.
   - If the resolver does not obtain resource information, the resolver continues to request A records, but searches the next domain in the sequence, which is `user`.

If the name server at 10.43.25.200 fails to respond to all of the queries, the system-wide total request count and the total failure count for this name server are incremented by 4. The network operator notification function would consider this name server to be unresponsive (because it experienced a 100% failure rate), but the autonomic quiescing of unresponsive name servers function would not (because less than 10 requests were directed to the name server). If this query were repeated three times within the 30-second monitoring interval and the total request count and total failure count for this name server was 12 and not 4, then the autonomic quiescing of unresponsive name servers function would consider this name server to be unresponsive because more than 10 queries were directed to the name server.

Optimizing the UNRESPONSIVETHRESHOLD value for your network

Once every monitoring interval, the resolver calculates the percentage of queries to a name server that failed in the previous 30-seconds or 5 minutes, and then compares this percentage to the threshold value that you set in the UNRESPONSIVETHRESHOLD statement to determine whether that DNS name server is unresponsive. If the resolver sends a query to a name server multiple times and the name server does not respond to multiple queries, each query is considered to be a unique failure to respond. When you specify the UNRESPONSIVETHRESHOLD value, consider the following factors that have an impact on the effectiveness of your setting:

- If you specify a small percentage for this value, an excessive number of operator notifications might occur. Short network disruptions that occur during the 30-second or 5-minute monitoring interval might result in some undeliverable resolver queries or name server responses, and a low threshold value might cause the resolver to alert the operator, and possibly stop using the name server, unnecessarily.
- If you specify a large percentage for this value, persistent issues with the network or the name server might be undetected even though a significant portion of resolver queries are not being processed by the name server.
- The setting on the RESOLVERTIMEOUT statement in the TCPIP.DATA file also affects the value that you should specify for the UNRESPONSIVETHRESHOLD setting. If you set a very short timeout value, even slight network disruptions might cause name server responses to be delayed longer than the amount of time specified by the RESOLVERTIMEOUT value. These delays are considered to be non-responses from the name server, which might cause unnecessary messages to be generated for this name server. A less aggressive (higher) percentage setting for the UNRESPONSIVETHRESHOLD value might be warranted in such a situation.
- The settings of the RESOLVERUDPRETRIES, SEARCH, and NAMESERVER statements in the TCPIP.DATA file can also contribute to high numbers of apparent failures on the part of the name server. See “Examples of resolver monitoring of DNS name servers” on page 766 for information about how these settings can influence the statistics that are collected by the resolver.

Guideline: When you set the optimal threshold by determining the error rate for a given name server, determine the error rate before you activate the autonomic quiescing of unresponsive name servers function.

One strategy that you can use to select the most optimal threshold value is to start with the default setting, which is 25%, and determine how many network operator messages are issued, if any, during normal operation of the network.

- If your network is operating in an acceptable manner (for example, no performance issues are detected and no host name or IP address resolutions delays are detected), examine the number of network operator alerts that are generated by the resolver:
  - If the number of network operator messages is zero or insignificant, leave the setting at the default value, or even decrease the threshold value slightly.
  - If the number of network operator messages is excessive, which suggests that a lot of false negative conditions were detected by the resolver, increase the threshold setting until the number of messages that is generated is appropriate for your network.
- If the name server is now responsive, but the failure rate is just slightly below the threshold value, the name server will probably become unresponsive again with a minor disruption in the network. If your network is currently operating in a satisfactory manner, consider increasing the threshold setting so that the resolver issues EZZ9308E messages only when your network conditions change significantly. Use

the statistics that are displayed when message EZZ9309I is issued to modify the threshold setting to a more optimal value.

- If your network is experiencing performance issues that resolver delays might be contributing to (for example, unexplained application delays), consider decreasing the responsiveness threshold setting to determine whether issues with the name servers are being detected by the resolver but are not being reported as unresponsive. If this lower threshold value causes the resolver to generate network operator messages that identify name servers that are unresponsive and that are impacting network operations, consider using this lower value for normal operations to provide more timely identification of name server issues.

A second strategy that you can use to select the most optimal threshold value is to start with the lowest threshold setting, which is 1%. If your name servers are failing to respond to a small percentage of the overall resolver queries that are being sent, the resolver generates EZZ9308E messages. At 5-minute intervals, the resolver also generates EZZ9310I messages, which indicate the percentage of failures for the most recent 5-minute sliding window. Use the EZZ9310I messages to determine the highest failure rate during normal operation of the network, and then set the threshold value to that rate, or to a value slightly above that rate. For example, if the highest failure percentage displayed on the EZZ9310I messages is 4%, set the threshold value to 5% for your network. This value ensures that the resolver considers name servers to be unresponsive only when they experience a failure rate that is greater than the rate that typically occurs in your network.

Steps for modifying the UNRESPONSIVETHRESHOLD value

You can modify your threshold value and restart or refresh the resolver.

Before you begin

You must have already created a resolver setup file; see “Steps for creating a resolver setup file” on page 742 for instructions.

Procedure

Perform the following steps to modify the UNRESPONSIVETHRESHOLD value.

1. Specify the UNRESPONSIVETHRESHOLD value that you want to use in the resolver setup file:
   - To disable the monitoring function, specify UNRESPONSIVETHRESHOLD(0).
   - To notify the network operator, specify UNRESPONSIVETHRESHOLD(*percentage*), where *percentage* is a value in the range 1 - 100. Do not specify the AUTOQUIESCE operand on the UNRESPONSIVETHRESHOLD statement.
   - To automatically quiesce unresponsive name servers, specify UNRESPONSIVETHRESHOLD(*percentage*,AUTOQUIESCE), where *percentage* is a value in the range 1 - 100.

     Rule: If you do not specify the UNRESPONSIVETHRESHOLD statement in the resolver setup file, the resolver notifies the network operator using the default value, which is 25%.
2. Perform one of the following steps:
   - If the resolver is not active, start the resolver.
   - If the resolver is currently active, issue the MODIFY RESOLVER,REFRESH,SETUP=*setup_file_name* command to cause the resolver to use the new threshold setting.

     If you did not specify the AUTOQUIESCE operand on the UNRESPONSIVETHRESHOLD statement, the new *percentage* value is used at the end of the next sliding 5-minute interval to determine name server responsiveness.

     If you specified the AUTOQUIESCE operand on the UNRESPONSIVETHRESHOLD statement, the new *percentage* value is used at the end of the next 30-second interval to determine name server responsiveness.

Results

You know you are done when the correct UNRESPONSIVETHRESHOLD value and AUTOQUIESCE values are displayed after you issue the START command or the correct values are displayed in the MODIFY RESOLVER,REFRESH command output.

Extension Mechanisms for DNS standards and the resolver

The resolver can use UDP protocols to more efficiently obtain resource information when it uses the Extension Mechanisms for DNS (EDNS0) standards. Before these standards existed, UDP responses from a name server were limited to 512 bytes. If a large number of resource records appear on a DNS response message, more than 512 bytes might be required to return all the response data to the resolver. IPv6 resource records are larger than IPv4 resource records, so fewer IPv6 resource records are needed to reach the 512 byte limitation, but the limitation can be reached even with just IPv4 resource records. EDNS0 support permits the resolver to accept DNS messages, using UDP protocols, of greater than 512 bytes, if the name server that is providing the response message also supports EDNS0.

- If the name server does not support EDNS0, these larger responses are truncated to fit within 512 bytes of UDP packet data, and the resolver resends the request using TCP protocols to acquire the entire response message.
- If the name server does support EDNS0, the resolver accepts up to 3072 bytes of DNS response message data in a single UDP packet.

You do not need to configure support for EDNS0 standards. If the resolver is not certain of the EDNS0 capability of a name server, the resolver does not use EDNS0 for that name server and does not indicate on any query that it sends to the name server that it supports EDNS0 processing. After the resolver dynamically determines that a name server supports EDNS0 processing, the resolver modifies the DNS requests that are sent to the name server to use EDNS0.

The resolver dynamically attempts to determine the EDNS0 capabilities of a name server the first time that the resolver receives a truncated UDP response from the name server; the resolver sends the same query to the name server and includes an indication that it supports EDNS0 processing, which is called an EDNS0 probe. The resolver determines whether the name server supports EDNS0 based on the following possible results from the EDNS0 probe:

- If the name server response to the EDNS0 probe indicates that the name server also supports EDNS0, the resolver uses EDNS0 on all subsequent queries to that name server.
- If the name server response to the EDNS0 probe indicates that the name server does not support EDNS0, the resolver does not use EDNS0 on subsequent queries to that name server.
- If the resolver does not receive a response before the timeout period expires, it does not use EDNS0 on subsequent queries to that name server.

  No response can indicate any of the following conditions:

  – The RESOLVERTIMEOUT value was too small for the larger UDP packet to be received in time
  – The name server ignored the EDNS0 probe query
  – A network router along the path from the name server to the resolver is discarding UDP packets greater than 512 bytes even though the name server and the resolver both support EDNS0

Because the resolver cannot determine why a timeout occurs, it does not use EDNS0 to that name server for a minimum five-minute interval. At the end of that interval, if appropriate, the resolver sends another EDNS0 probe query to determine whether the name server now supports EDNS0. To fully gain EDNS0 performance benefits, you should choose a resolver timeout value that is long enough to allow larger UDP packets to arrive.

The resolver periodically verifies that the name server does not support EDNS0, even if the response to the EDNS0 probe explicitly indicated that the name server does not support EDNS0. The periodic EDNS0 probe processing allows the resolver to dynamically discover that the capabilities of the name server have changed, although the rediscovery period might take some time. You can use the MODIFY RESOLVER,REFRESH command to cause the resolver to rediscover the capabilities of the name

servers more quickly. For more information about the MODIFY RESOLVER,REFRESH command, see z/OS Communications Server: IP System Administrator's Commands.

To verify whether a name server supports EDNS0, use the **dig** command with the +bufsize= option to force **dig** to send an OPT RR record on the request. If the name server supports EDNS0, it responds with its own OPT RR record on the response.

If you have upgraded a name server to support EDNS0, you can issue the MODIFY RESOLVER,REFRESH command to force the resolver to dynamically determine name server capability. The resolver can then use EDNS0 support to accept DNS messages of greater than 512 bytes, using the less costly UDP protocol, which results in improved DNS and resolver performance.

Resolver configuration files

Understanding the resolver search orders used in native MVS and z/OS UNIX environments is key to setting up your system properly.

The resolver can use available name servers, local definitions, or a combination of both, to process API resolver requests. Figure 105 on page 772 shows how local definitions can be specified and searched for when needed.

Use the Trace Resolver facility to determine what TCPIP.DATA values are being used by the resolver and where they were read from. For information about dynamically starting the trace or using the Resolver CTRACE function to collect Trace Resolver output, see z/OS Communications Server: IP Diagnosis Guide. After the trace is active, issue the Netstat HOME/-h command to display the values. You can issue a `Ping` of a host name from TSO and from the z/OS UNIX shell to show the activity to the resolver cache and to any DNS servers that might be configured.

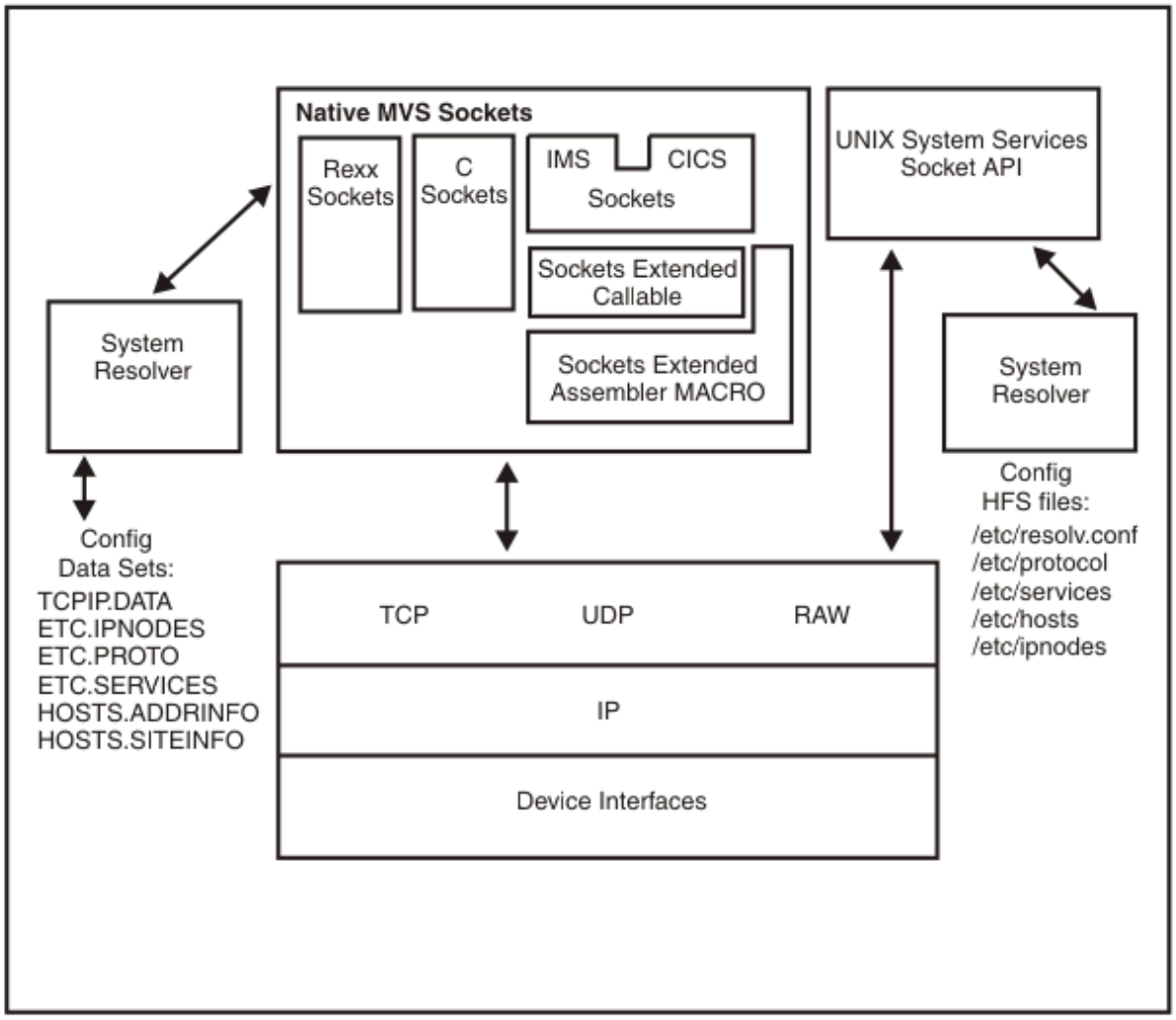


Figure 105. Resolver related configuration files in z/OS UNIX and native MVS environments

Table 36 on page 772 shows the complete set of local definition possibilities available to the resolver.

Table 36. Local definitions available to resolver

File type description	**APIs affected**	**Candidate files**
Base resolver configuration files	All APIs	1. GLOBALTCPIPDATA 2. RESOLVER_CONFIG environment variable 3. /etc/resolv.conf 4. SYSTCPD DD-name 5. *userid*.TCPIP.DATA 6. *jobname*.TCPIP.DATA 7. SYS1.TCPPARMS(TCPDATA) 8. DEFAULTTCPIPDATA 9. TCPIP.TCPIP.DATA

Table 36. Local definitions available to resolver (continued)

File type description	APIs affected	Candidate files
Translate tables	All APIs	1. X_XLATE environment variable 2. *userid*.STANDARD.TCPXLBIN 3. *jobname*.STANDARD.TCPXLBIN 4. *hlq*.STANDARD.TCPXLBIN 5. Resolver-provided translate table, member STANDARD in SEZATCPX
Local host tables	endhostent endnetent getaddrinfo gethostbyaddr gethostbyname gethostent GetHostNumber GetHostResol GetHostString getnameinfo getnetbyaddr getnetbyname getnetent IsLocalHost Resolve sethostent setnetent	1. X_SITE environment variable 2. X_ADDR environment variable 3. /etc/hosts 4. *userid*.HOSTS.xxxxINFO 5. *jobname*.HOSTS.xxxxINFO 6. *hlq*.HOSTS.xxxxINFO 7. GLOBALIPNODES 8. RESOLVER_IPNODES environment variable 9. *userid*.ETC.IPNODES 10. *jobname*.ETC.IPNODES 11. *hlq*.ETC.IPNODES 12. DEFAULTIPNODES 13. /etc/ipnodes
Protocol information	endprotoent getprotobyname getprotobynumber getprotoent setprotoent	1. /etc/protocol 2. *userid*.ETC.PROTO 3. *jobname*.ETC.PROTO 4. *hlq*.ETC.PROTO
Services information	endservent getaddrinfo getnameinfo getservbyname getservbyport getservent setservent	1. /etc/services 2. SERVICES DD-name 3. *userid*.ETC.SERVICES 4. *jobname*.ETC.SERVICES 5. *hlq*.ETC.SERVICES
Host alias table	getaddrinfo gethostbyname	HOSTALIASES environment variable

The actual search order of the candidate files varies depending on the type of API that is used and the resolver setup. The search orders are explained in more detail in "Search orders used in the z/OS UNIX environment" on page 777 and "Search orders used in the native MVS environment" on page 781. Because the resolver runs in the address space of the application, the resolver accesses the candidate files from the application address space.

Use the Trace Resolver facility to obtain information about an application's search order. Trace Resolver output provides a caller API value that determines which search order is used. For information about

dynamically starting the trace or using the Resolver CTRACE function to collect Trace Resolver output, see z/OS Communications Server: IP Diagnosis Guide.

The following caller API values indicate the z/OS UNIX environment search order is used:

- Language Environment C Sockets
- UNIX System Services [z/OS UNIX System Services (z/OS UNIX) callable services]

The following caller API values indicate the native MVS environment search order is used:

- TCP/IP C Sockets
- TCP/IP Pascal Sockets
- TCP/IP Rexx Sockets
- TCP/IP Sockets Extended

The following commands are some examples of Communications Server TSO commands that use the native MVS search order:

- DIG
- FTP (batch only)

  Rule: Batch FTP jobs use //SYSTCPD if it is specified. If //SYSTCPD is not specified, then the z/OS UNIX search order is used.
- LPR
- NETSTAT
- NSLOOKUP
- PING
- REXEC
- RPCINFO
- RSH
- TRACERTE

The following commands are some examples of Communications Server UNIX commands that use the z/OS UNIX search order:

- **dig**
- **dnsdomainname**
- **domainname**
- **ftp**

  Rule: The TSO FTP command also uses the z/OS UNIX search order.
- **host**
- **hostname**
- **netstat**
- **nslookup**
- **ping**
- **rexec**
- **rpcinfo**
- **snmp**
- **traceroute**

The following applications are some examples of Communications Server applications that use the native MVS search order:

- CICS Listener

- LPD
- Miscellaneous server
- PORTMAP
- RSHD
- TN3270E Telnet server

The following applications are some examples of Communications Server applications that use the z/OS UNIX search order:

- CSSMTP
- FTP
- SNMP agent
- z/OS UNIX OPORTMAP
- z/OS UNIX OREXECD
- z/OS UNIX ORSHD

z/OS XL C/C++ environment variables for configuration files

A z/OS XL C/C++ environment variable is an identifier used like a variable in a program. In Table 36 on page 772, the following environment variables appear:

HOSTALIASES
The host aliases data set, file, or ddname.

RESOLVER_CONFIG
The resolver configuration data set, file, or ddname. The RESOLVER_CONFIG environment variable is used by TCP/IP to include the name of an MVS data set or z/OS UNIX file in the search order for TCPIP.DATA.

RESOLVER_IPNODES
The IPNODES data set, file, or ddname.

X_SITE and X_ADDR
The HOSTS.SITEINFO and HOSTS.ADDRINFO data sets or ddnames created by the MAKESITE TSO command. The X_SITE environment variable influences how gethostbyname() resolves the network address of the specified host name. The X_ADDR environment variable is used by some TCP/IP functions, such as getnetbyaddr(), to include the name of an MVS data set or z/OS UNIX file in the search order for the HOSTS.ADDRINFO data set.

X_XLATE
The ASCII-EBCDIC translate table data set or ddname created by the CONVXLAT TSO command. The X_XLATE environment variable is used by TCP/IP to include the name of an MVS data set or z/OS UNIX file in the search order for the STANDARD.TCPXLBIN data set.

Other environment variables that can be explicitly set by the resolver include the following variables:

LOCALDOMAIN
Defines the domain origin. When this environment variable is set, it overrides any setting for DOMAIN, DOMAINORIGIN, or SEARCH found in TCPIP.DATA

RESOLVER_TRACE
Defines the data set, file, or ddname into which the resolver trace output is written.

MESSAGECASE
Determines whether messages are translated to all uppercase characters before being sent to the console.

Setting z/OS XL C/C++ environment variables

The method that you use to set z/OS XL C/C++ environment variables so that a z/OS XL C/C++ UNIX application is able to retrieve the value depends on whether you start the application from the z/OS shell or from JCL. The following general methods apply to all environmental variables; the examples are specific to the RESOLVER_CONFIG environmental variable.

Setting z/OS XL C/C++ environment variables from the z/OS shell

If you are starting the z/OS XL C/C++ UNIX application from the z/OS shell, you must set the environment variables using the **export** shell command. You can specify either the file name or the data set name, as shown in the following two examples for setting the RESOLVER_CONFIG environment variable:

- To set the value of the RESOLVER_CONFIG environment variable to the file name /etc/tcpa.data, specify the following **export** command:

```
export RESOLVER_CONFIG=/etc/tcpa.data
```

- To set the RESOLVER_CONFIG environment variable to the data set TCPA.MYFILE(TCPDATA), specify the following **export** command.

```
export RESOLVER_CONFIG="//'TCPA.MYFILE(TCPDATA)'"
```

  Rule: You must put single quotation marks around the data set name. If you do not, your user ID will be added as a prefix to the data set name when the resolver tries to open the file.

Setting z/OS XL C/C++ environment variables from JCL

If you are starting the z/OS XL C/C++ UNIX application from JCL, you must code the environment variable as a parameter in the JCL of the application, as shown in the following examples:

- This example shows setting the RESOLVER_CONFIG environment variable to a z/OS UNIX file:

```
//OSNMPD    PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)',
//       'ENVAR("RESOLVER_CONFIG=/etc/tcpa.data")/-d 0')
⋮
```

- This example shows setting the RESOLVER_CONFIG environment variable to an MVS partitioned data set:

```
//OSNMPD    PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)',
//  'ENVAR("RESOLVER_CONFIG=//''TCPA.MYFILE(TCPDATA)''")/-d 0')
⋮
```

- This example shows setting the RESOLVER_CONFIG environment variable to a DD card:

```
//OSNMPD PROC
//*
//* Procedure for running the SNMP agent
//*
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM=('POSIX(ON) ALL31(ON)',
// 'ENVAR("RESOLVER_CONFIG=DD:TCPDATA")/-d 0')
//TCPDATA DD DSN=TCPA.MYFILE(TCPDATA),DISP=SHR
⋮
```

Tip: You can also specify a ddname as //DD:*ddname*, as shown in the following example:

```
⋮
// 'ENVAR("RESOLVER_CONFIG=//DD:TCPDATA")/-d 0')
```

For more information about specifying a ddname, see z/OS XL C/C++ Programming Guide.

- This example shows how you can set environment variables using the STDENV DD statement. The environment variables will be read from the location specified by the STDENV DD statement. The advantage of using STDENV is that it eliminates any concern about the maximum number of characters that can be specified on the PARM= parameter of the EXEC JCL statement. Perform the following steps to set environment variables using the STDENV DD statement:

  1. Create the MVS data set that will contain the environment variables with a variable length record format (RECFM=V/VB).

     This example uses an MVS data set TCPA.ENVAR(OSNMPD).

     Rule: Do not use fixed length record format (RECFM=F/FB) to allocate the MVS data set because fixed length pads the environment variables with blanks.

  2. Edit the MVS data set so that it contains the environment variables.

     The following example shows setting the RESOLVER_CONFIG environment variable to a file:

     ```
     RESOLVER_CONFIG=/etc/tcpa.data
     ```

     The following example shows setting the RESOLVER_CONFIG environment variable to an MVS data set:

     ```
     RESOLVER_CONFIG=//'TCPA.MYFILE(TCPDATA)'
     ```

  3. Update the application's JCL to pick up the TCPIP.DATA information from the specified RESOLVER_CONFIG environment variable.

     ```
     //OSNMPD PROC
     //*
     //* Procedure for running the SNMP agent
     //*
     //OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
     // PARM=('POSIX(ON) ALL31(ON)',
     // 'ENVAR("_CEE_ENVFILE=DD:STDENV")/-d 0')
     //STDENV DD DSN=TCPA.ENVAR(OSNMPD),DISP=SHR
     ```

  For more information about specifying a list of environment variables using the _CEE_ENVFILE environment variable, see z/OS XL C/C++ Programming Guide.

Search orders used in the z/OS UNIX environment

This information describes the search orders used in the z/OS UNIX environment for the different file types shown in Table 36 on page 772. The z/OS UNIX socket functions use various types of TCP/IP data sets and files. They include:

- Base resolver configuration files
- Translate tables
- Local host tables
- Protocol information
- Services information
- Host alias table

The particular file or table chosen can be either an MVS data set or z/OS UNIX file, depending on the resolver configuration settings and the presence of given files on the system.

Note: A program's first resolver service request initializes the resolver definitions that will be used for all resolver requests. For long running programs, the definitions can be modified by use of the MODIFY

REFRESH operator command. For command usage and syntax, see z/OS Communications Server: IP System Administrator's Commands.

Base resolver configuration files

The base resolver configuration file contains TCPIP.DATA statements. In addition to resolver directives, it is referenced to determine, among other things, the data set prefix (DATASETPREFIX statement's value) to be used when trying to access some of the configuration files.

The search order used to access the base resolver configuration file is as follows:

1. GLOBALTCPIPDATA

   If defined, the resolver GLOBALTCPIPDATA setup statement value is used. For a description of the GLOBALTCPIPDATA statement, see "The resolver and the global TCPIP.DATA file" on page 741.

   The search continues for an additional configuration file. The search ends with the next file found.
2. The value of the environment variable RESOLVER_CONFIG

   The value of the environment variable is used. This search will fail if the file does not exist or is allocated exclusively elsewhere.
3. /etc/resolv.conf
4. //SYSTCPD DD card

   The data set allocated to the ddname SYSTCPD is used. In the z/OS UNIX environment, a child process does not have access to the SYSTCPD DD. This is because the SYSTCPD allocation is not inherited from the parent process over the fork() or exec function calls.
5. *userid*.TCPIP.DATA

   userid is the user ID that is associated with the current security environment (address space or task/thread)
6. SYS1.TCPPARMS(TCPDATA)
7. DEFAULTTCPIPDATA

   If defined, the resolver DEFAULTTCPIPDATA setup statement value is used. For a description of the DEFAULTTCPIPDATA statement, see "The resolver and the global TCPIP.DATA file" on page 741.
8. TCPIP.TCPIP.DATA

Any TCPIP.DATA statements that have not been found will have their default values, if any, assigned.

Translate tables

The translate tables (EBCDIC-to-ASCII and ASCII-to-EBCDIC) are referenced to determine the translate data sets to be used.

The search order that is used to access this configuration file is as follows. The search order ends at the first file found:

1. The value of the environment variable X_XLATE

   The value of the environment variable is the name of the translate table that is produced by the CONVXLAT TSO command.
2. *userid*.STANDARD.TCPXLBIN

   userid is the user ID that is associated with the current security environment (address space or task/thread).
3. *hlq*.STANDARD.TCPXLBIN

   hlq represents the value of the DATASETPREFIX statement that is specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.
4. If no table is found, the resolver uses a hardcoded default table that is identical to the STANDARD member in the SEZATCPX data set.

Tip: You can use a JCL DD statement to preallocate the STANDARD.TCPXLBIN data set and prevent the resolver from dynamically allocating it. The name on the DD statement can be any valid ddname. Preallocation prevents the dynamic allocation messages (for example, IEF237I and IEF285I) from being written to the job output log.

Local host tables

By default, resolver first attempts to use any configured domain name servers for resolution requests. If the resolution request cannot be satisfied, local host tables are used. Resolver behavior is controlled by the following factors:

- TCPIP.DATA statements

  The TCPIP.DATA resolver statements define if and how domain name servers are to be used. The LOOKUP TCPIP.DATA statement can also be used to control how domain name servers and local host tables are used. For more information on TCPIP.DATA statements, see z/OS Communications Server: IP Configuration Reference.
- How your application is written and compiled

  If your application program uses the TCP/IP-provided C/C++ API and the XL C/C++ RESOLVE_VIA_LOOKUP symbol was defined, only local host tables will be used. For information on the use of the RESOLVE_VIA_LOOKUP symbol, see z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference and z/OS XL C/C++ Programming Guide.

The local host table supplies sitename information for, as one example, resolving hostnames to host or network addresses. The local host table can also supply address information, for example, for resolving addresses to hostname or network names. There are different search orders used for selecting the local host table for these different purposes. The search order to use is based on certain resolver setup statements, the type of API invocation, and possibly the type of host address (IPv4 versus IPv6) being requested or being resolved.

IPv4-unique search order for sitename information

The resolver uses the IPv4-unique search order for sitename information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the:

- getaddrinfo API is attempting to locate an IPv4 address.
- gethostbyname, sethostent, gethostent, or endhostent API is invoked.

If the COMMONSEARCH statement is specified, see “IPv6/common search order” on page 780, where the resolver can use IPNODES to locate sitenames.

The resolver uses the IPv4-unique search order for sitename information unconditionally for getnetbyname API calls.

The IPv4-unique search order for sitename information is as follows. The search ends at the first file found:

1. The value of the environment variable X_SITE

   The value of the environment variable is the name of the MVS data set that contains the sitename information. This data set is created by the TSO MAKESITE command.
2. /etc/hosts
3. *userid*.HOSTS.SITEINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).
4. *hlq*.HOSTS.SITEINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv4-unique search order for address information

The resolver uses the IPv4-unique search order for address information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the getnameinfo API is attempting to resolve an IPv4 address or the gethostbyaddr API is invoked. If the COMMONSEARCH statement is specified, see "IPv6/common search order" on page 780, where the resolver can use IPNODES to locate IPv4 and IPv6 addresses.

The resolver uses the IPv4-unique search order for address information unconditionally for the setnetent, getnetent, endnetent, or getnetbyaddr APIs.

The IPv4-unique search order for address information is as follows. The search ends at the first file found:

1. The value of the environment variable X_ADDR

   The value of the environment variable is the name of the MVS data set that contains the address information. This data set is created by the TSO MAKESITE command.
2. /etc/hosts
3. *userid*.HOSTS.ADDRINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).
4. *hlq*.HOSTS.ADDRINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv6/common search order

The resolver uses the IPv6/common search order when it determines that any of the following conditions exist:

- The resolver setup statement COMMONSEARCH is specified (to have the resolver use IPNODES to locate IPv4 addresses, IPv6 addresses, and sitenames), and the getaddrinfo, gethostbyname, getnameinfo, gethostbyaddr, sethostent, gethostent, or endhostent APIs are invoked.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getaddrinfo API is attempting to locate an IPv6 address.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getnameinfo API is attempting to resolve an IPv6 address.

Note: The IPv6/common search order is never used for the following API socket calls:

- getnetbyname
- getnetbyaddr
- setnetent
- getnetent
- endnetent

The IPv6/common search order is as follows. The search ends at the first file found:

1. GLOBALIPNODES value

   If defined, the resolver GLOBALIPNODES setup statement value is used. For a description of the GLOBALIPNODES statement, see "The resolver setup file" on page 738.
2. The value of the environment variable RESOLVER_IPNODES
3. *userid*.ETC.IPNODES

   userid is the user ID that is associated with the current security environment (address space or task/thread).
4. *hlq*.ETC.IPNODES

hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

5. DEFAULTIPNODES

   If defined, the resolver DEFAULTIPNODES setup statement value is used. For a description of the DEFAULTIPNODES statement, see "The resolver setup file" on page 738.

6. /etc/ipnodes

Protocol information

The protocol information supplies protocol related information for the socket calls listed in Table 36 on page 772.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. /etc/protocol
2. *userid*.ETC.PROTO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

3. *hlq*.ETC.PROTO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Services information

The services information supplies the service information for the socket calls listed in Table 36 on page 772.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. /etc/services
2. *userid*.ETC.SERVICES

   userid is the user ID that is associated with the current security environment (address space or task/thread).

3. *hlq*.ETC.SERVICES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Host alias table

The host alias table supplies hostname alias information for the socket calls listed in Table 36 on page 772. The format of the alias information is the alias name, followed by a space, followed by the fully qualified domain name that corresponds to the alias name. The domain name is written without a trailing dot, and the alias name cannot contain dots. The search order used to access this configuration file consists only of the value of the environment variable HOSTALIASES.

Search orders used in the native MVS environment

The native MVS environment socket functions use various type of TCP/IP data sets, including:

- Base resolver configuration files
- Translate tables
- Local host tables
- Protocol information
- Services information

The particular file or table chosen depends on the resolver configuration settings and the presence of given files on the system.

Note: A program's first resolver service request initializes the resolver definitions that will be used for all resolver requests. For long running programs, the definitions can be modified by use of the MODIFY REFRESH operator command. For command usage and syntax, see z/OS Communications Server: IP System Administrator's Commands.

Base resolver configuration files

The base resolver configuration file contains TCPIP.DATA statements. In addition to resolver directives, it is referenced to determine, among other things, the data set prefix (DATASETPREFIX statement's value) to be used when trying to access some of the configuration files.

The search order used to access the base resolver configuration file is as follows:

1. GLOBALTCPIPDATA.

   If defined, the resolver GLOBALTCPIPDATA setup statement value is used. For a description of the GLOBALTCPIPDATA statement, see "The resolver and the global TCPIP.DATA file" on page 741.

   The search continues for an additional configuration file. The search ends with the next file found.

2. //SYSTCPD DD card

   The data set allocated to the ddname SYSTCPD is used.

   Rule: Because TCPIP.DATA statements might need to be read and used multiple times by the resolver, the FREE=CLOSE JCL parameter should not be used when allocating SYSTCPD. To allow TCPIP.DATA statements to be changed while still allocated for long running programs, consider using a member of an MVS partitioned data set instead of an MVS sequential data set. For these long running applications, the resolver MODIFY REFRESH command should then be used to indicate that TCPIP.DATA statements have been changed.

3. *userid/jobname*.TCPIP.DATA

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

4. SYS1.TCPPARMS(TCPDATA)
5. DEFAULTTCPIPDATA

   If defined, the resolver DEFAULTTCPIPDATA setup statement value is used. For a description of the DEFAULTTCPIPDATA statement, see "The resolver and the global TCPIP.DATA file" on page 741.

6. TCPIP.TCPIP.DATA

Translate tables

The translate tables are referenced to determine the translate data sets to be used.

The search order that is used to access this configuration file is as follows. The search order ends at the first file found:

1. *userid/jobname*.STANDARD.TCPXLBIN

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name that is specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

2. *hlq*.STANDARD.TCPXLBIN

hlq represents the value of the DATASETPREFIX statement that is specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

3. If no table is found, the resolver uses a hardcoded default table that is identical to the STANDARD member in the SEZATCPX data set.

Tip: You can use a JCL DD statement to preallocate the STANDARD.TCPXLBIN data set and prevent the resolver from dynamically allocating it. The name on the DD statement can be any valid ddname. Preallocation prevents the dynamic allocation messages (for example, IEF237I and IEF285I) from being written to the job output log.

Local host tables

By default, resolver first attempts to use any configured domain name servers for resolution requests. If the resolution request cannot be satisfied, local host tables are used. Resolver behavior is controlled by the following factors:

- TCPIP.DATA statements

  The TCPIP.DATA resolver statements define if and how domain name servers are to be used. The LOOKUP TCPIP.DATA statement can also be used to control how domain name servers and local host tables are used. For more information on TCPIP.DATA statements, see z/OS Communications Server: IP Configuration Reference.

- How your application is written and compiled

  If your application program uses the TCP/IP-provided C/C++ API and the RESOLVE_VIA_LOOKUP symbol was defined, only local host tables will be used. For information on the use of the RESOLVE_VIA_LOOKUP symbol, see z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference.

The local host table supplies sitename information for, as one example, resolving hostnames to host or network addresses. The local host table can also supply address information, for example, for resolving addresses to hostname or network names. There are different search orders used for selecting the local host table for these different purposes. The search order to use is based on certain resolver setup statements, the type of API invocation, and possibly the type of host address (IPv4 versus IPv6) being requested or being resolved.

IPv4-unique search order for sitename information

The resolver uses the IPv4-unique search order for sitename information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the:

- getaddrinfo API is attempting to locate an IPv4 address.
- gethostbyname, GetHostNumber, GetHostResol, IsLocalHost, Resolve, sethostent, gethostent, or endhostent API is invoked.

If the COMMONSEARCH statement is specified, see "IPv6/common search order" on page 784, where the resolver can use IPNODES to locate sitenames.

The resolver uses the IPv4-unique search order for sitename information unconditionally for getnetbyname API calls.

The IPv4-unique search order for sitename information is as follows. The search ends at the first file found:

1. *userid/jobname*.HOSTS.SITEINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

2. *hlq*.HOSTS.SITEINFO

hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv4-unique search order for address information

The resolver uses the IPv4-unique search order for address information when the resolver setup statement NOCOMMONSEARCH is specified (or left to default), and either the getnameinfo API is attempting to resolve an IPv4 address or the gethostbyaddr or GetHostString API is invoked. If the COMMONSEARCH statement is specified, see "IPv6/common search order" on page 784, where the resolver can use IPNODES to locate IPv4 and IPv6 addresses.

The resolver uses the IPv4-unique search order for address information unconditionally for the setnetent, getnetent, endnetent, or getnetbyaddr APIs.

The IPv4-unique search order for address information is as follows. The search ends at the first file found:

1. *userid/jobname*.HOSTS.ADDRINFO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.
2. *hlq*.HOSTS.ADDRINFO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

IPv6/common search order

The resolver uses the IPv6/common search order when it determines that any of the following conditions exist:

- The resolver setup statement COMMONSEARCH is specified (to have the resolver use IPNODES to locate IPv4 addresses, IPv6 addresses, and sitenames), and the getaddrinfo, gethostbyname, getnameinfo, gethostbyaddr, GetHostNumber, GetHostResol, GetHostString, IsLocalHost, Resolve, sethostent, gethostent, or endhostent APIs are invoked.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getaddrinfo API is attempting to locate an IPv6 address.
- The resolver setup statement NOCOMMONSEARCH is specified (or left to default), and the getnameinfo or Resolve API is attempting to resolve an IPv6 address.

Note: The IPv6/common search order is never used for the following API socket calls:

- getnetbyname
- getnetbyaddr
- setnetent
- getnetent
- endnetent

The IPv6/common search order is as follows. The search ends at the first file found:

1. GLOBALIPNODES value

   If defined, the resolver GLOBALIPNODES setup statement value is used. For a description of the GLOBALIPNODES statement, see "The resolver setup file" on page 738.
2. *userid/jobname*.ETC.IPNODES

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

3. *hlq*.ETC.IPNODES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); otherwise, *hlq* is TCPIP by default.

4. DEFAULTIPNODES

   If defined, the resolver DEFAULTIPNODES setup statement value is used. For a description of the DEFAULTIPNODES statement, see "The resolver setup file" on page 738.

5. /etc/ipnodes

Protocol information

The protocol information supplies protocol related information for the socket calls listed in Table 36 on page 772.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. *userid/jobname*.ETC.PROTO

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

2. *hlq*.ETC.PROTO

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Services information

The services information supplies service information for the socket calls listed in Table 36 on page 772.

The search order used to access this configuration file is as follows. The search ends at the first file found:

1. //SERVICES DD card

   The data set allocated to the ddname SERVICES is used.

2. *userid/jobname*.ETC.SERVICES

   userid is the user ID that is associated with the current security environment (address space or task/thread).

   jobname is the name specified on the JOB JCL statement for batch jobs or the procedure name for a started procedure.

3. *hlq*.ETC.SERVICES

   hlq represents the value of the DATASETPREFIX statement specified in the base resolver configuration file (if found); Otherwise, *hlq* is TCPIP by default.

Chapter 14. Policy-based networking

Businesses typically define goals for network behavior in human terms. Network implementations provide a variety of application-transparent controls for priority treatment of traffic, bandwidth management, security, and control of network behavior. The link between the high-level business goals and network implementations is defined as policy-based networking and is provided by policies. Policies are usually defined in a centralized repository and are accessed by nodes that need to make policy decisions (Policy Decision Point, or PDP) or implement such decisions (Policy Enforcement Point, or PEP).

Policy types and infrastructure overview

To implement networking policies for your users, you must use the z/OS Communications Server policy infrastructure. You can use the policy types supported by the Policy Agent for any of the following purposes:

- Policy-based routing (See "Policy-based routing" on page 349)
- Quality of service (See Chapter 15, "Quality of service," on page 831)
- Intrusion detection services (See Chapter 16, "Intrusion detection services," on page 851)
- IP filtering, and manual and dynamic virtual private network (VPN) tunnels, collectively referred to as IPSec policies (See Chapter 17, "IP security," on page 883)
- Application Transparent Transport Layer Security (AT-TLS, see Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121)
- zERT policy-based enforcement (ZERT, see "Defining a zERT enforcement policy" on page 176)

For more information about the policy types, see "Policy types" on page 801.

Based on the policy types that you want to implement, you must configure and start one or more policy infrastructure components:

- TCP/IP stack

  TCP/IP stacks implement most of the policy types. You need to start one or more stacks per logical partition (LPAR).

- Syslog daemon (syslogd)

  Syslogd acts as the central message logging facility for z/OS UNIX applications. Syslogd is not specific to the policy infrastructure, but the policy infrastructure depends on syslogd to provide a central logging facility to maintain an audit trail. If you do not start syslogd, messages are lost. You should start one syslog daemon per LPAR.

- Policy Agent

  You must start Policy Agent to install and maintain policies in the TCP/IP stacks in an LPAR. You need one Policy Agent per LPAR.

- Traffic regulation management daemon (TRMD)

  TRMD formats and sends policy-related messages to your syslog daemon. You need one TRMD per TCP/IP stack in an LPAR.

- Internet Key Exchange daemon (IKED)

  IKED is used for negotiating and setting up dynamic VPN tunnels. If you are not using dynamic VPN tunnels, you do not need to start IKED; otherwise, you need one IKED per LPAR.

- Network security services daemon (NSSD)

  NSSD can be used as the central certificate and key server for z/OS IKE daemons, or as a network security server for selected non-z/OS platforms. NSSD can be used independently of any z/OS

© Copyright IBM Corp. 2000, 2022

networking policies, but is an element of the overall z/OS networking policy infrastructure. Typically, you do not need an NSSD on every LPAR; one NSSD per sysplex is more likely.

- Defense Manager daemon (DMD)

  DMD provides support for short-term defensive filters. You can use DMD without defining any IPSec filter policies, but typically you use DMD in addition to IPSec filter policy. You need one DMD per LPAR.

- Network service level agreement performance monitor 2 (NSLAPM2)

  NSLAPM2 is an SNMP subagent that provides QoS metrics through MIB variables. You need one NSLAPM2 per TCP/IP stack in an LPAR.

For more information about syslogd, see "Configuring the syslog daemon" on page 213. For more information about the other policy infrastructure components, see "Policy infrastructure components" on page 793.

To determine the policy infrastructure components that you need to start based on which policy types you are implementing, see Table 37 on page 788.

Table 37. Policy components needed per policy type

Policy type	**Component**							
	One or more instances per LPAR	**One instance per LPAR**					**One instance per TCP/IP stack in an LPAR**	
	TCP/IP stack	**Policy Agent**	**syslogd**	**IKED**	**NSSD**	**DMD**	**NSLAPM2**	**TRMD**
QoS	Required	Required	Required				Optional	
IDS	Required	Required	Required					Required
AT-TLS	Required	Required	Required					
IPSec filters	Required	Required	Required			Optional		Required
IPSec VPNs	Required	Required	Required	Optional (dynamic VPNs)	Optional (central key and certificate server)			Required
Policy-based routing	Required	Required	Required					
zERT policy-based enforcement	Required	Required	Required					Required

You can use the IBM Configuration Assistant for z/OS Communications Server for assistance with setting up and configuring security, JCL procedures, and configuration files for the following policy infrastructure components:

- Policy Agent, including policy definition files for QoS, IDS, AT-TLS, IPSec, policy-based routing, and zERT policy-based enforcement
- IKED

- NSSD
- DMD

Configuration files and policy definition files

To operate correctly, the policy infrastructure depends on various configuration files and policy definitions files.

The IBM Configuration Assistant for z/OS Communications Server enables flat-file configuration of all supported policy types for z/OS. The IBM Configuration Assistant for z/OS Communications Server is an optional GUI-based tool that provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the Policy Agent files.

The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

Table 38 on page 789 lists the policy-related configuration and definition files, whether the files can be created using the IBM Configuration Assistant for z/OS Communications Server, and the default location of the configuration files. You can manually edit all files listed in Table 38 on page 789.

Table 38. Configuration files and policy definition files

Configuration file or policy definition file	Can be created by the IBM Configuration Assistant for z/OS Communications Server?	Default z/OS location
Configuration files		
Policy agent	Yes	/etc/pagent.conf
syslogd	No	/etc/syslog.conf
DMD	Yes	/etc/security/dmd.conf
IKED	Yes	/etc/security/iked.conf
NSSD	Yes	/etc/security/nssd.conf
Policy definition files		
QoS	Yes	None
IDS	Yes	None
AT-TLS	Yes	None
IPSec	Yes	None
Policy-based routing	Yes	None
zERT policy-based enforcement	Yes	None

Managing changes to configuration files and policy definition files

Typically, you use a change management procedure when you install a set of new or changed policies, or when you modify the configuration files for the Policy Agent or other policy-related applications. A structure for storing and maintaining the configuration files and policy definition files that are related to the general policy infrastructure is provided by the IBM Configuration Assistant for z/OS Communications Server, and you can use it to manually create and edit the configuration files and policy definition files. The structure is based on three location levels:

- Staging

  The staging level is the location in which you manually edit your definitions, or into which you upload definitions from the IBM Configuration Assistant for z/OS Communications Server.

- Production

  The production level is the location to which you copy your staging definitions when you are ready to put your changes into production. This is the location from which your Policy Agent and other policy infrastructure components read their definitions.

- Backup and recovery

  The backup and recovery level is the location to which you copy your existing production definitions before you copy your staging definitions to your production location. If the changes are in error, you can back out the changes by copying your original production definitions from the recovery location to the production location and restarting your policy infrastructure components.

Storing configuration files and policy definition files

You can store configuration files and policy definition files in the z/OS UNIX file system or in traditional MVS data sets.

Some definitions are shared by all TCP/IP stacks on a z/OS image, and some definitions are specific to individual TCP/IP stacks on a z/OS image.

For example, if you are using the z/OS UNIX file system to store your configuration files and policy definition files, a directory structure for z/OS image SYSA with TCP/IP stacks TCPIP1 and TCPIP2 might look as follows:

- Staging

  Image-wide
  /etc/tcpip/POLTRANS/SYSA/

  TCPIP1
  /etc/tcpip/POLTRANS/SYSA/TCPIP1/

  TCPIP2
  /etc/tcpip/POLTRANS/SYSA/TCPIP2/

- Production

  Image-wide
  /etc/tcpip/POLPROD/SYSA/

  TCPIP1
  /etc/tcpip/POLPROD/SYSA/TCPIP1/

  TCPIP2
  /etc/tcpip/POLPROD/SYSA/TCPIP2/

- Backup and recovery

  Image-wide
  /etc/tcpip/POLBACK/SYSA/

  TCPIP1
  /etc/tcpip/POLBACK/SYSA/TCPIP1/

  TCPIP2
  /etc/tcpip/POLBACK/SYSA/TCPIP2/

Similarly, you can use MVS partitioned data set (PDS) or partitioned data set extended (PDSE) libraries to store the configuration files and policy definition files as follows:

- Staging

  Image-wide
  hlq.TCPPARMS.POLTRANS.SYSA

TCPIP1
hlq.TCPPARMS.POLTRANS.SYSA.TCPIP1

TCPIP2
hlq.TCPPARMS.POLTRANS.SYSA.TCPIP2

- Production

Image-wide
hlq.TCPPARMS.POLPROD.SYSA

TCPIP1
hlq.TCPPARMS.POLPROD.SYSA.TCPIP1

TCPIP2
hlq.TCPPARMS.POLPROD.SYSA.TCPIP2

- Backup and recovery

Image-wide
hlq.TCPPARMS.POLBACK.SYSA

TCPIP1
hlq.TCPPARMS.POLBACK.SYSA.TCPIP1

TCPIP2
hlq.TCPPARMS.POLBACK.SYSA.TCPIP2

You can maintain a number of members in each of these libraries, using your own naming convention or the following suggested naming convention:

- LPAR-wide configuration files

  CONF (configuration file)
- Stack-specific configuration and policy definition files
  - TLSPOL (AT-TLS policy definitions)
  - IDSPOL (IDS policy definitions)
  - IPSPOL (IPSec policy definitions)
  - QOSPOL (QoS policy definitions)
  - PBRPOL (Policy-based routing policy definitions)
  - ZERTPOL (zERT policy-based enforcement policy definitions)

The following example shows the IKE daemon configuration file and the IPSec policy definitions for stack TCPIP1 as members of PDS or PDSE libraries:

```
USER.LPAR1.POLTRANS.SYSA.IKED(CONF)
USER.LPAR1.POLTRANS.SYSA.TCPIP1(IPSPOL)
```

Steps for managing policy changes

Place your new policy flat file in the staging library, back up your current production policy flat file, and then replace the production copy with the new copy.

Procedure

As shown in Figure 106 on page 792, perform the following steps to activate changes to your policies:

1. Transfer the policy flat file to the staging library. *hlq*.TCPPARMS.POLTRANS.SYSA.TCPIP1(*xyz*)
2. Back up the current production policy flat file.
   Copy *hlq*.TCPPARMS.POLPROD.SYSA.TCPIP1(*xyz*) to *hlq*.TCPPARMS.POLBACK.SYSA.TCPIP1(*xyz*)
3. Copy the new policy flat file into production.
   Copy *hlq*.TCPPARMS.POLTRANS.SYSA.TCPIP1(*xyz*) to *hlq*.TCPPARMS.POLPROD.SYSA.TCPIP1(*xyz*)

Results

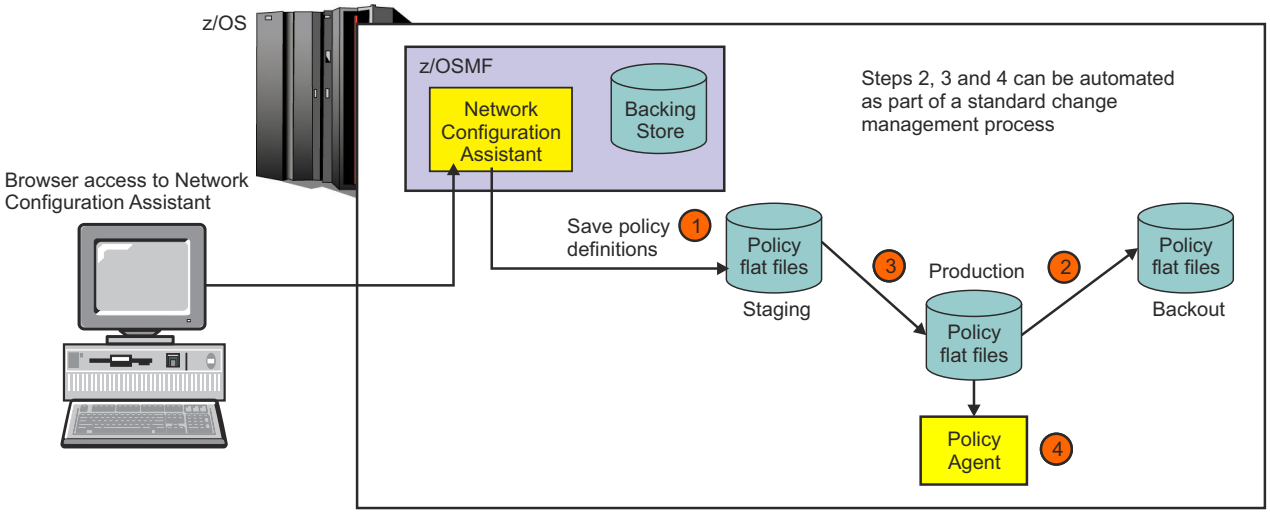


Figure 106. Activating changes to your policies

Policy Agent reads the definitions from the production location, as shown by step 4 in Figure 106 on page 792.

Tips:

- If the changes you make are in error, you can back out the changes by copying your original production definitions from the recovery location to the production location and restarting your policy infrastructure components.
- In your procedure to move objects from POLTRANS to POLPROD, you might have to include a utility step to change location references from POLTRANS to POLPROD.

  Some configuration files might include references to other configuration file locations or policy definition file locations. Those locations are initially the POLTRANS location.
- You can override names suggested by the IBM Configuration Assistant for z/OS Communications Server and choose to use other names.

  If you are using the IBM Configuration Assistant for z/OS Communications Server, you are prompted for the base location for a z/OS image. For each configuration file, policy definition file, sample RACF job, sample JCL procedure, and so on, the IBM Configuration Assistant for z/OS Communications Server suggests a name to serve as the file name in the z/OS UNIX file system directories, or as the member name in the PDS or PDSE libraries.
- Specifying an EBCDIC code page that matches your default country settings makes it easier to browse the files with ISPF.

  Most policy-related configuration files and all the policy definition files support an optional Codepage parameter that you can use to indicate the EBCDIC code page in which the contents are encoded. When you use the IBM Configuration Assistant for z/OS Communications Server, you specify the EBCDIC code page when you define the base location for a z/OS image. The IBM Configuration Assistant for z/OS Communications Server transfers configuration files and policy definitions files to that z/OS image using the specified code page and includes the Codepage option in the files. The following example shows the initial section of a Policy Agent configuration file that uses the Codepage option:

```
## Configuration Assistant, 2008.11.25 09:33:12
## Pagent Configuration for system MVS098
## Target Codepage = IBM-278
TcpImage TCPCS 'USER1.TCPPARMS.MVS098.TCPCS(STKPAG)' FLUSH NOPURGE 360
Codepage  IBM-278
## Loglevel 127         ## default 31
```

Policy infrastructure components

Based on the policy types that you want to implement, you must configure and start one or more policy infrastructure components.

For more information about syslogd, see “Configuring the syslog daemon” on page 213.

TCP/IP stack

Most policy types are implemented by the TCP/IP stack. IPSec dynamic VPNs are implemented by the IKE daemon.

As packets are sent or received, they are matched against policies of the appropriate type as needed. In general, policy processing occurs at the following layers in the stack:

- Application layer - AT-TLS
- IP layer - IPSec filtering, IDS, and policy-based routing
- Transport layer - IDS , QoS and zERT policy-based enforcement (ZERT)

Some types of IDS checks are also performed at the IP layer in the stack.

When a matching policy is found, it is implemented against the packet. Depending on the policy type and packet contents, this results in a wide variety of actions. For example, the packet might be discarded, processed according to its priority, or have its routing changed. For information about the policy action statements and the kinds of processing that can be applied for each policy type, see the Policy Agent and policy applications information in z/OS Communications Server: IP Configuration Reference.

Policy Agent

Policy Agent can act in any of several roles, and provides various services, such as managing dependent components of the policy infrastructure.

The term *Policy Agent* refers to any roles or services provided by Policy Agent. The terms *policy server* and *policy client* refer to those specific roles, and the term *import requester* refers to the TCP/IP address discovery service.

Policy Agent roles

The Policy Agent runs in the z/OS environment and can act in any of several roles, depending on configuration options:

- The Policy Agent can act as a self-contained Policy Decision Point (PDP) on a single system, installing policies in one or more z/OS Communications Server stacks on that z/OS image, as shown by system SYSA in Figure 107 on page 794.
- The Policy Agent can act as a policy client, retrieving remote policies from the policy server, as shown by system SYSB in Figure 107 on page 794. Each stack in a Common INET (CINET) environment that is configured to the Policy Agent acts as a separate policy client.
- The Policy Agent can act as a centralized policy server, providing PDP services for one or more remote policy clients, as shown by system SYSC in Figure 107 on page 794.
- A single Policy Agent can act as a policy client or a policy server, but not both.

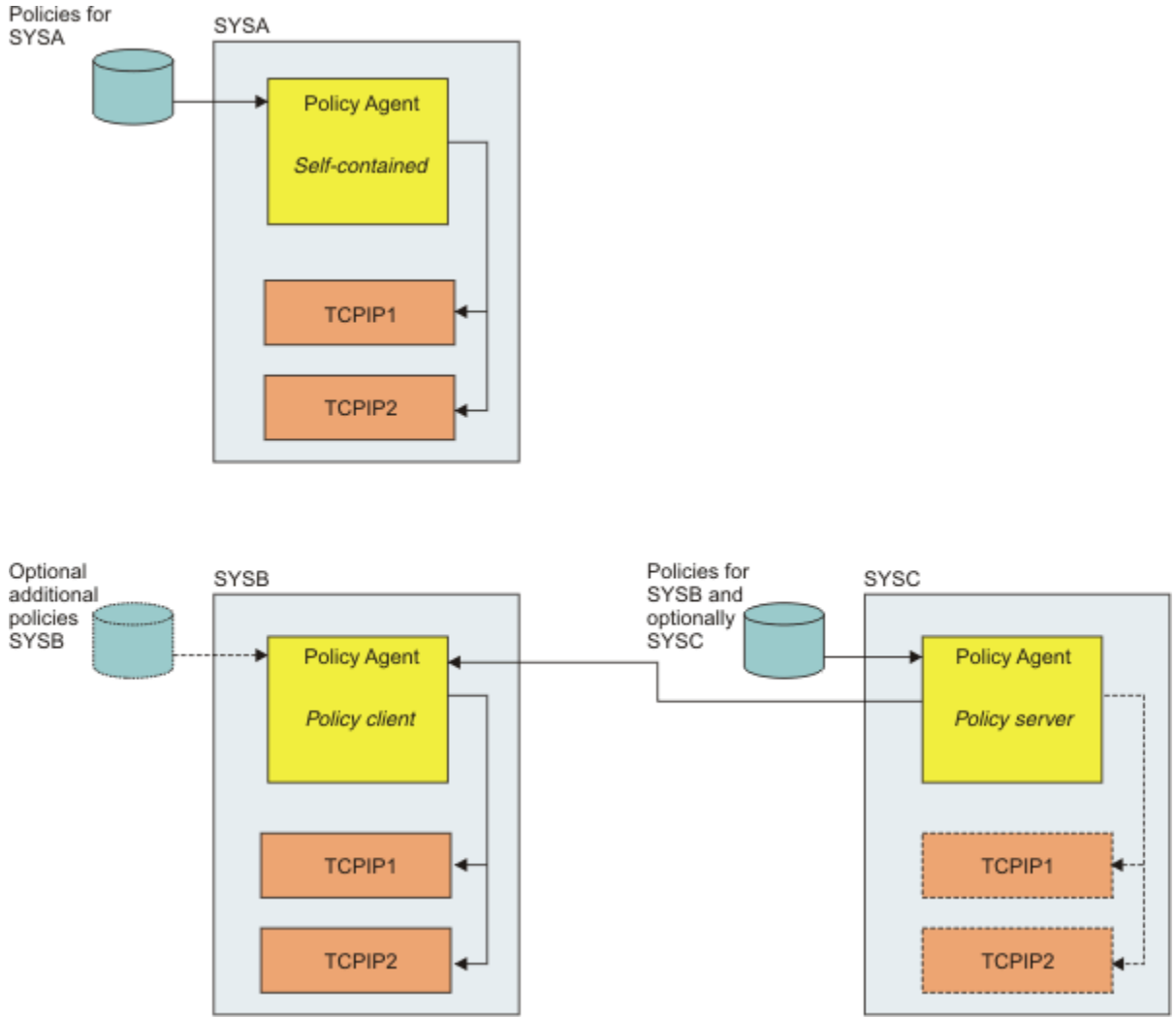


Figure 107. Policy Agent roles

Policy Agent services

Policy Agent provides the following services:

- Reads and parses policy definitions, and provides those policies to one or more TCP/IP stacks or policy clients. The Policy Agent actively participates in maintaining the policies by monitoring the policy files, detecting changes, and informing the TCP/IP stacks or policy clients about any changes.
- Imports TCP/IP interface information through a passive service. The main user of this service is the IBM Configuration Assistant for z/OS Communications Server (Network Configuration Assistant) IPSec technology. This function imports TCP/IP interface information from a running stack into Network Configuration Assistant for use as IPSec stack symbols and is not related to the TCP/IP profile import function of Network Configuration Assistant.

  For this request, the Policy Agent retrieves information from the TCP/IP stack. The import requester can retrieve this information and use it as configuration data.
- Starts, monitors, stops, and restarts dependent components of the policy infrastructure. You can instruct Policy Agent to manage the following components:
  - DMD
  - IKED
  - NSSD
  - syslogd

– TRMD per stack on a system

Policy Agent policies

Policies can be defined in several different ways.

Table 39 on page 795 shows the format you can use for different policy types.

Table 39. Policy formats

Policy type	Text file format [1]	LDAP format
QoS	Yes	Yes [2]
IDS	Yes	Yes [2]
AT-TLS	Yes	No
IPSec	Yes	No
Policy-based routing	Yes	No
zERT policy-based enforcement	Yes	No

[1] The IBM Configuration Assistant for z/OS Communications Server builds policy definitions only in text file format.

[2] The format of the LDAP schema for IDS and QoS policies was stabilized in z/OS V1R2. Only IDS and QoS policies supported by that release are supported when you are using an LDAP server to store IDS and QoS policies. For information about defining policies on an LDAP server, see Appendix F, "Using an LDAP server for policy definitions," on page 1377.

When acting as the Policy Decision Point (PDP) for a single system, Policy Agent can read policy definitions from local configuration files, a central repository that uses the Lightweight Directory Access Protocol (LDAP), or both. The Policy Agent also installs policies in one or more z/OS Communications Server stacks. It can be used to replace existing policies or to update them as necessary.

When acting as a policy server, Policy Agent also acts as a PDP for the local system, and thus can read policies from local configuration files or an LDAP server and install them in local stacks. However, it also reads policies from local configuration files on behalf of policy clients. These policies are retrieved by policy clients, but are not installed in the local stacks on the policy server.

Restriction: Dynamic monitoring for file updates using the `-i` startup option is not supported for files read on behalf of policy clients.

When acting as a policy client, Policy Agent retrieves remote policies from the policy server, and can also use local policies from configuration files or an LDAP server. The choice of local or remote policies can be made separately for each supported policy type (QoS, IDS, IPSec, Routing, AT-TLS or ZERT). The policy client informs the policy server of its local capabilities, so that the policy server can perform appropriate parsing of the policies. For example, the policy client might not support the IPSec 3DES encryption algorithm, so the policy server needs to fail IPSec policies that specify 3DES, even if the policy server itself does support 3DES.

If the policy client and policy server are at different release levels, you must be careful when defining policies on the policy server.

- If the policy client is at a higher release level than that of the policy server, you can define policies using the syntax and semantics of only the lower-level policy server. You cannot use the capabilities that exist only in the higher release level, such as new statements, parameters, or parameter values. Error checking might also be unavailable for use; rules or restrictions might be added to or removed from the higher release level.
- If the policy server is at a higher release level than that of the policy client, you cannot define policies using syntax and semantics that are available only in the higher release level. The policy server cannot

parse such policies on behalf of a policy client at a lower release level, so the policies are reported as containing errors.

- If the policy server is at a lower release level than that of the policy client, any polices using syntax and semantics that were removed from the higher release level are still parsed and returned from the policy server to the policy client.

As a general rule, configure policies based on the target system, not on the system where the policies are defined.

For a table of statements, parameters, parameter values, and rules or restrictions that are valid only for certain release levels, see z/OS Communications Server: IP Configuration Reference.

The policy client can be configured with a backup as well as a primary policy server. The policy client continually tries the connection to the primary policy server (and the connection to the backup if a backup is configured) using the connection retry values configured on the ServerConnection statement, until a connection is successfully established.

For more information, see the following topics:

- "Policy-based routing" on page 349
- Chapter 15, "Quality of service," on page 831
- Chapter 16, "Intrusion detection services," on page 851
- Chapter 17, "IP security," on page 883
- Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121
- "Defining a zERT enforcement policy" on page 176

Restriction: You cannot define AT-TLS policies, IPSec policies, routing policies, or ZERT on LDAP servers.

Tip: Policies defined on an LDAP server use the configuration files and mechanisms provided by the LDAP server product. The definition of the elements of policies is known as the *schema*. z/OS Communications Server provides the schema definition for policies that may be defined on an LDAP server in a set of sample files. The sample files are provided in LDAP protocol version 3 format (see "LDAP sample files" on page 1387 for the names of these samples). These sample files must be installed on the LDAP server as the schema definition. Policy Agent uses the z/OS Integrated Security Server LDAP Client library to communicate with an LDAP server. See z/OS IBM Tivoli Directory Server Administration and Use for z/OS for more information about LDAP. A copy of the LDAP definition files that define the policy definitions for LDAP is available in z/OS Communications Server: IP Configuration Reference.

Local policies are defined in Policy Agent configuration files, in the LDAP server, or both. Remote policies are defined in Policy Agent configuration files on the policy server. Policies from configuration files and the LDAP server are combined into a single list. This requires unique policy object names per type (QoS, IDS, IPSec, Routing, AT-TLS, and ZERT). On a policy client, policies for a given type are retrieved either locally or remotely, but not both.

For policies defined on the LDAP server, the distinguished name (DN) must be unique, but the user-friendly name does not have to be unique (although it should be). The Policy Agent appends a unique suffix if it is required to make LDAP user-friendly names unique within the scope of policies defined on the LDAP server. When policies from a configuration file are combined with LDAP-defined policies, the LDAP user-friendly names must be unique with respect to the names defined in the configuration file. Any policy objects of the same type (that is, QoS or IDS) with duplicate names at this point are discarded by the Policy Agent and an error is reported.

Import services

The IBM Configuration Assistant for z/OS Communications Server can request TCP/IP interface information be imported from a running TCP/IP stack for use as stack symbols by IPSec technology. When the Policy Agent provides this import service, the IBM Configuration Assistant for z/OS Communications Server is acting as an *import requester*. When the IBM Configuration Assistant for z/OS Communications Server is acting as an import requester, the required input values are configured on request panels for discovery import (the Discover Stack Local Addresses panel):

- Host connection IP address and port for the Policy Agent
- Host connection user name and password to identify resources that this user can access
- Indication of whether a secure connection (SSL) should be used. For more details, see "Step 6: Configure Policy Agent for import services" on page 818.

Restrictions:

- The import requester's port value must be the same as the port value defined on the ServicesConnection statement.
- The import requester's TLS/SSL security configuration must match the security configuration for the Policy Agent provider.

For detailed configuration information, see "Step 6: Configure Policy Agent for import services" on page 818.

Additional QoS services

The Policy Agent supports QoS functions other than reading and installing policies, such as sysplex distributor policy performance monitoring, and mapping IPv4 Type of Service (ToS) byte or IPv6 Traffic Class values to outbound interface and virtual LAN (VLAN) user priorities. The QoS specific Policy Agent functions are further described in "QoS-specific Policy Agent functions" on page 833.

Policy API

A Policy API (PAPI) is provided to allow access to policy information by external user applications. The PAPI interface can be used by policy performance monitoring applications to retrieve policy performance data. For more information on PAPI, see z/OS Communications Server: IP Programmer's Guide and Reference.

Traffic regulation management daemon

The TRMD application provides the ability for IDS, IPSec, and ZERT messages to be logged to syslogd. A separate instance of TRMD must be started for each TCP/IP stack.

IKE daemon

IPSec services include IP filtering and support for manual and dynamic VPN tunnels. The Internet Key Exchange (IKE) daemon works with the stack to provide IPSec support. IPSec policy can be defined for IP filtering (including manual VPN tunnels), key exchange, and dynamic VPN tunnels.

Network security services daemon

A network security services daemon (NSSD) provides network security services for one or more security disciplines, including IPSec. For the IPSec discipline, these services include the IPSec certificate service and the IPSec remote management service. For more information, see Chapter 18, "Network security services," on page 1083.

Tip: You do not define policies for NSSD.

Defense Manager daemon

The Defense Manager daemon (DMD) provides defensive filters, which are IP filter rules to discard packets that are separate from IP security filters, and which are typically installed for a short duration (for example, 30 minutes) to block a specific attack or a pattern of attacks. For more information, see Chapter 19, "Defensive filtering," on page 1107.

Tip: You do not define policies for DMD.

SNMP Network SLAPM2 subagent

The z/OS CS Network SLAPM2 subagent (nslapm2) allows network administrators to retrieve data and determine if the current set of Network SLAPM2 policy definitions are performing as needed or if adjustments need to be made. The Network SLAPM2 subagent supports the Network Service Level Agreement Performance Monitor (NETWORK-SLAPM2) MIB. See usr/lpp/tcpip/samples/slapm2.mi2 for more information about the Network SLAPM2 MIB.

Sample policy infrastructure

Figure 108 on page 799 shows the Policy Agent in a sample policy infrastructure.

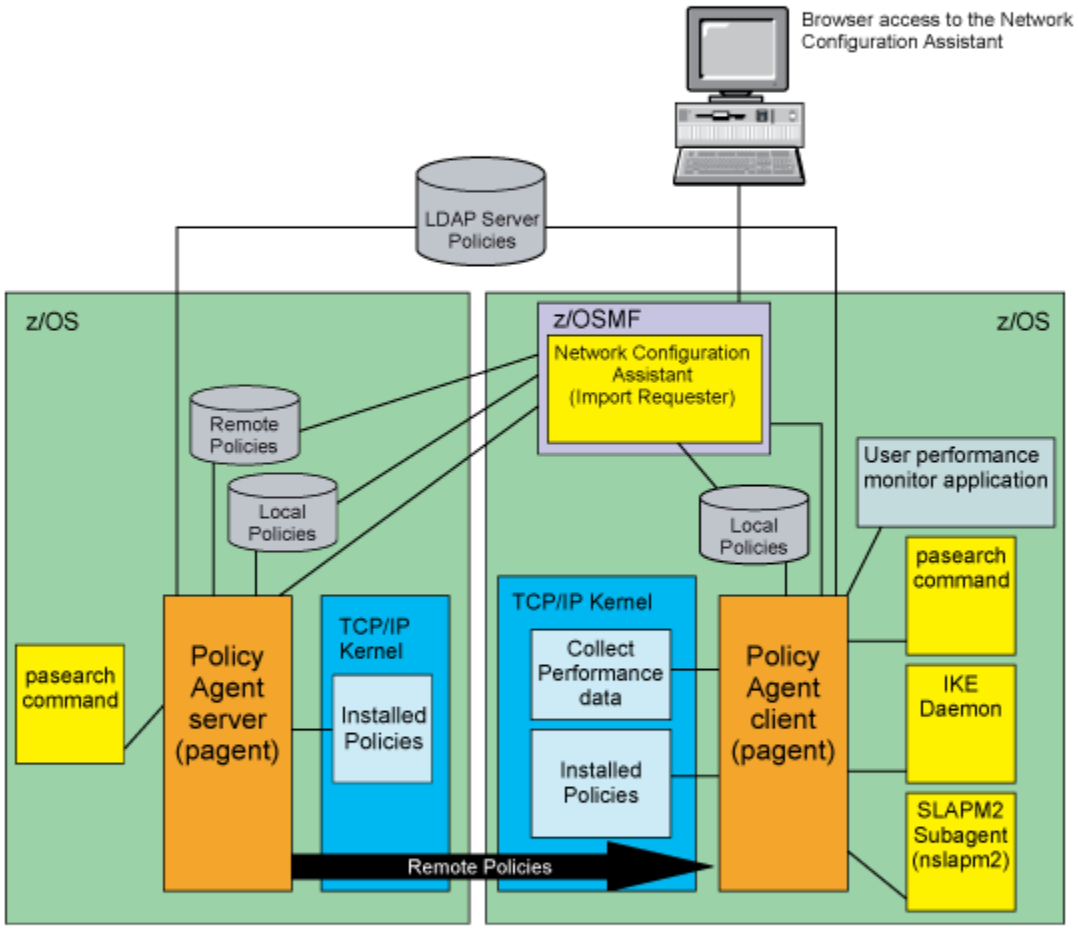


Figure 108. Sample policy infrastructure

The IBM Configuration Assistant for z/OS Communications Server performs the following functions:

- GUI to build the configuration flat-files for IPSec, AT-TLS, IDS, Routing, ZERT, and QoS policies.
- Import requester to request TCP/IP interface information import services from the Policy Agent, for use as stack symbols by IPSec technology. For details, see "Import services" on page 796.

Policy sample files

A set of sample files is shipped with z/OS CS that provides several functions. The first sample file provides an example of policy definitions in a Policy Agent configuration file.

/usr/lpp/tcpip/samples/pagent.conf
This file contains overall policy definition rules, syntax and semantics for defining policies in a configuration file, and examples of such policy definitions.

The next set of sample files provide sample IPSec policy definitions.

/usr/lpp/tcpip/samples/pagent_CommonIPSec.conf
This file contains sample common IPSec policy definitions. These can be referenced and reused by multiple stack-specific IPSec configuration files.

/usr/lpp/tcpip/samples/pagent_IPSec.conf
This file contains sample stack-specific IPSec policy definitions. Some of these refer to common definitions in /usr/lpp/tcpip/samples/pagent_CommonIPSec.conf.

The following file provides sample AT-TLS policy definitions.

/usr/lpp/tcpip/samples/pagent_TTLS.conf
This file contains sample AT-TLS policy definitions. These definitions can either be in a common or stack-specific AT-TLS file. If these definitions are in a common AT-TLS file, they can be referenced and reused by multiple stack-specific AT-TLS configuration files. If these definitions are in a stack-specific AT-TLS file, they are used only by that specific stack.

The following file provides sample IDS policy definitions.

/usr/lpp/tcpip/samples/pagent_IDS.conf
This file contains sample IDS policy definitions. These definitions can either be in a common or stack-specific IDS file. If these definitions are in a common IDS file, they can be referenced and reused by multiple stack-specific IDS configuration files. If these definitions are in a stack-specific IDS file, they are used by only that specific stack.

The following file provides sample policy-based routing policy definitions.

/usr/lpp/tcpip/samples/pagent_Routing.conf
This file contains sample policy-based routing policy definitions. These definitions can either be in a common or stack-specific routing file. If these definitions are in a common routing file, they can be referenced and reused by multiple stack-specific routing configuration files. If these definitions are in a stack-specific routing file, they are used by only that specific stack.

The following file provides sample zERT policy-based enforcement policy definitions.

/usr/lpp/tcpip/samples/pagent_ZERT.conf
This file contains sample zERT policy-based enforcement policy definitions. These definitions can only be in a stack-specific ZERT file, hence, they are used only by that specific stack.

The following files include sample C applications that can be used to develop policy performance monitoring applications.

/usr/lpp/tcpip/samples/pagent/README
This file contains instructions for compiling and running the following sample C applications.

/usr/lpp/tcpip/samples/pagent/pCollector.c
This file is a sample C application (pCollector) that uses the Policy API (PAPI) interfaces to access policy performance data. It can be used as the base for an application that provides near real-time policy performance monitoring.

/usr/lpp/tcpip/samples/pagent/pCollector.h
This file is a header file for the pCollector sample application.

/usr/lpp/tcpip/samples/pagent/pLogReader.c
This file is a sample C application (pLogReader) that reads the policy performance log file to access policy performance data. It can be used as the base for an application that provides offline policy performance monitoring.

This documentation refers to Version 1 through Version 4 when defining policies.

- Version 1 refers to policy definitions defined with the ServicePolicyRules and ServiceCategories statements or LDAP objects.
- Version 2 through Version 4 refer to policy definitions defined with other policy statements or LDAP objects.
- The primary difference between Version 2 and Version 3 is in the definition of the LDAP schema.
- Version 3 is used with configuration file IPSec and AT-TLS policies.
- Version 4 is used with configuration file IDS, Routing, and ZERT policies.

For information about LDAP samples and schema definition files, see Appendix F, "Using an LDAP server for policy definitions," on page 1377.

Policy types

The Policy Agent supports the following types of policies. Each policy type is referred to as a discipline.

- Quality of service (QoS) policies
  - Differentiated Services (DS) policies
  - Integrated Services (RSVP) policies
  - Sysplex distributor (SD) policies
- Intrusion detection services (IDS) policies
  - Scan policies
  - Attack policies
  - Traffic Regulation policies
- IP security (IPSec) policies
  - IP filtering policies
  - Key exchange policies
  - Local dynamic VPN policies
- Application Transparent Transport Layer Security (AT-TLS) policies
- Policy-based routing (Routing) policies
- zERT policy-based enforcement (ZERT) policies

For information about how IPv6 affects the Policy Agent and which types of policies support IPv6, see z/OS Communications Server: IPv6 Network and Application Design Guide.

These policy types are defined using different policy schemas. They use a common rule, but have separate conditions and actions. None of the different policy types can be mixed in a given policy object. All policy rules can contain time-related information that indicates when the policy rule should be considered active or inactive.

For the QoS, IDS, Routing, , AT-TLS and ZERT types, active policy rules are installed in the TCP/IP stack, so they can be applied as traffic filters, while inactive policy rules exist only in the Policy Agent. For the IPSec type, both active and inactive IP filtering policies are installed in the TCP/IP stack. However, only manual VPN tunnels that are active as a result of a time condition are installed in the stack. For the Routing policy type, active route tables are installed in the stack, while inactive route tables exist only in the Policy Agent. Configured route tables are active when they are referenced by an active Routing rule and its associated Routing action.

The Policy Agent supports all of the previously mentioned policy types, installing them into one or more TCP/IP stacks as configured. However, policies to be retrieved by policy clients are not installed in any stacks on the policy server.

QoS policy

Policy conditions consist of a variety of selection criteria that act as traffic filters. Traffic can be filtered based on source and destination IP addresses, source and destination ports, protocol, inbound and outbound interfaces, application name, application specific data or application priority. Only packets that match the filter criteria are selected to receive the accompanying action. Policy rules can refer to several policy actions, but only one policy action is executed per policy scope. A given policy action may be referred to by several policy rules.

The type of policy defined is in general controlled by the policy scope value defined for the policy action. SD policies are an exception. SD policies are a subset of DS policies, so use the DS scope.

Although RSVP policies are installed into the TCP/IP stack, they are used only for collecting policy statistics. For policy use and limit enforcement, these policies are requested from the Policy Agent by the RSVP Agent, to apply to RSVP reservation requests from RSVP applications.

IDS policy

Policy conditions primarily determine the portion of IDS function that is being configured. A given IDS policy rule refers to a single IDS policy action. A given IDS policy action may be referred to by several policy rules of the same IDS type. See Chapter 16, “Intrusion detection services,” on page 851 for more details.

IPSec policy

Policy conditions consist of a variety of selection criteria that act as filters for IP filtering rules. Traffic can be filtered based on source and destination IP addresses, source and destination ports, protocol, direction, routing information, and security class. For other types of IPSec policies, policy conditions contain information about dynamic key exchange filters or dynamic VPN tunnels. For more details, see Chapter 17, “IP security,” on page 883.

IP filter rules and key exchange rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a simple rule, while one with more conditions is known as a complex rule. Complex IP filter rules and key exchange rules have their conditions evaluated according to Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. For details on CNF, see “Policy object model overview” on page 1377.

Complex IP filter rules (rules that contain groupings, or sets, of individual conditions) are split to produce multiple simple rules to be installed in the TCP/IP stack. The conditions in the IpFilterRule statement that can make a filter rule complex are:

- IpSourceAddr

  If multiple source addresses (or address ranges) are specified in a rule, the rule is considered complex. Multiple source addresses can be specified by referencing a set or group of addresses from the rule (IpSourceAddrGroupRef).

- IpDestAddr

  If multiple destination addresses (or address ranges) are specified in a rule, the rule is considered complex. Multiple destination addresses can be specified by referencing a set or group of addresses from the rule (IpDestAddrGroupRef).

- IpService

  If multiple IpService statements are specified in a rule, the rule is considered complex. Multiple IpService statements can be specified either inline or by referencing a group of IpService statements (IpServiceGroupRef).

- IpService Direction

  If the Direction parameter in an IpService statement is configured as bidirectional, the rule is considered complex.

Complex key exchange rules are split to produce multiple simple rules. The IKE daemon retrieves simple rules when necessary. The following conditions can make a complex key exchange rule:

- LocalSecurityEndpoint Location

  If multiple IP addresses (or address ranges) are specified in a local security endpoint, the associated key exchange rule is considered complex. You can specify multiple IP addresses by referencing a set or group of addresses from the local security endpoint (LocationGroupRef).

- RemoteSecurityEndpoint Location

  If multiple IP addresses (or address ranges) are specified in a remote security endpoint, the associated key exchange rule is considered complex. You can specify multiple IP addresses by referencing a set or group of addresses from the remote security endpoint (LocationGroupRef).

For more details on these IPSec policy configuration statements and parameters, see z/OS Communications Server: IP Configuration Reference.

The **pasearch** command displays IP filter rules and key exchange rules as complex rules, and not split as installed in the TCP/IP stack or retrieved by the IKE daemon.

For IP filter rules and key exchange rules, the condition level summaries are not applicable and are always displayed as all zeros.

AT-TLS policy

Policy conditions consist of a variety of selection criteria that act as filters for AT-TLS rules. Traffic can be filtered based on local addresses, remote addresses, local port range, remote port range, job name, user identification, and direction. For more details, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

AT-TLS policy rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a simple rule, while one with more conditions is known as a complex rule. Complex AT-TLS policy rules have their conditions evaluated according to Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. For details on CNF, see “Policy object model overview” on page 1377.

When AT-TLS rules are read and parsed, Policy Agent creates the rule as a complex rule. For example, consider the following TTLSRule statement:

```
TTLSRule ttlsRule1
{
  LocalAddrGroupRef    addrGroup1
  RemoteAddrGroupRef   addrGroup2
  LocalPortGroupRef    portGroup1
  RemotePortGroupRef   portGroup2
  Jobname              jobABC
  Userid               user1
  Direction            Outbound
  TTLSGroupActionRef   ttlsAction7
}

IpAddrGroup addrGroup1
{
  IpAddr
  {
    Addr 9.1.1.1
  }
  IpAddr
  {
    Addr 10.1.1.1
  }
}

IpAddrGroup addrGroup2
{
  IpAddr
  {
    Addr 200.1.1.1
  }
  IpAddr
  {
    Addr 201.1.1.1
```

```
  }
}

PortGroup portGroup1
{
  PortRange
  {
    Port 21
  }
  PortRange
  {
    Port 23
  }
}

PortGroup2
{
  PortRange
  {
    Port 10
  }
  PortRange
  {
    Port 15
  }
}
```

This rule is represented as a CNF rule with the following condition levels (levels are ANDed together):

- Level 1 = local address 9.1.1.1 OR local address 10.1.1.1
- Level 2 = remote address 200.1.1.1 OR remote address 201.1.1.1
- Level 3 = local port 21 OR local port 23
- Level 4 = remote port 10 OR remote port 15
- Level 5 = job name jobABC, user ID user1, direction outbound

The **pasearch** command displays the AT-TLS policy as complex rules.

Policy-based routing policy

Policy conditions consist of a variety of selection criteria that act as filters for policy-based routing (Routing) rules. Traffic can be filtered based on source addresses, destination addresses, source port range, destination port range, protocol, job name, security zone, and security label. For more details, see "Policy-based routing" on page 349.

Routing policy rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a simple rule, and one with more conditions is known as a complex rule. Complex routing policy rules have their conditions evaluated according to Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. For details about CNF, see "Policy object model overview" on page 1377.

When routing rules are read and parsed, Policy Agent creates the rule as a complex rule. For example, consider the following RoutingRule statement:

```
RoutingRule   rule1
{
  TrafficDescriptorGroupRef  tdGroup
  IpSourceAddrGroupRef       addrGroup
  RoutingActionRef           action1
}
TrafficDescriptor   td1
{
   SourcePortRange      1-5
   DestinationPortRange 10
   SecurityZone         zone1
   SecurityLabel        label1
   JobName              jobABC1
}
TrafficDescriptor   td2
{
   SourcePortRange      6-9
   DestinationPortRange 25
   SecurityZone         zone2
   SecurityLabel        label2
```

```
   JobName              jobABC2
}
TrafficDescriptorGroup  tdGroup
{
  TrafficDescriptorRef  td1
  TrafficDescriptorRef  td2
}
IpAddrGroup addrGroup
{
  IpAddr
  {
    Addr 9.1.1.1
  }
  IpAddr
  {
    Addr 10.1.1.1
  }
}
```

This rule is represented as a CNF rule with the following condition levels (levels are ANDed together):

- Level 1 = source address 9.1.1.1 OR source address 10.1.1.1
- Level 2 = (source port range 1-5 AND destination port range 10 AND job name jobABC1 AND security zone zone1 AND security label label1) OR (source port range 6-9 AND destination port range 25 AND job name jobABC2 AND security zone zone2 AND security label label2)

The **pasearch** command displays the Routing policy as a complex rule.

zERT policy-based enforcement (ZERT) policy

Policy conditions consist of a variety of selection criteria that act as filters for zERT policy-based enforcement (ZERT) rules. Connections can be filtered based on local addresses, remote addresses, local port range, remote port range, job name, user identification, and direction. A number of security protection attributes can also be used as filters. For more details, see “Defining a zERT enforcement policy” on page 176.

ZERT policy rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a simple rule, while one with more conditions is known as a complex rule. Complex ZERT policy rules have their conditions evaluated according to Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. For details about CNF, see “Policy object model overview” on page 1377. When ZERT rules are read and parsed, Policy Agent creates the rule as a complex rule. The pasearch command displays the ZERT policy as complex rules. For example, consider the following ZERTRule statement:

```
ZERTRule                         Rule1
{
   LocalAddrGroupRef              l~addrGroup
   RemoteAddrGroupRef             r~addrGroup
   ZERTActionRef                  act1
   SecurityProtocol               TLS
   ZERTTLSProtocolRef             tlsVersion
   ZERTSymmetricEncryptionRef     symEnc
   ZERTMessageAuthenticationRef   msgAuth
   ConnectionDescriptorGroupRef   connDescGrp
}

IpAddrGroup                   l~addrGroup
{
    IpAddr
    {
        Addr 9.1.1.1
    }
    IpAddr
    {
        Addr 10.1.1.1
    }
}
IpAddrGroup                   r~addrGroup
{
    IpAddr
    {
        Addr 200.1.1.1
    }
```

```
    IpAddr
    {
        Addr 201.1.1.1
    }
}

ZERTTLSProtocol tlsVersion
{
    TLSProtocol TLSv1.2
    TLSProtocol TLSv1.3
}
ZERTSymmetricEncryption        symEnc
{
    SymmetricEncryption        AES_CBC_128
    SymmetricEncryption        AES_CBC_256
    SymmetricEncryption        AES_GCM_256
}
ZERTMessageAuthentication      msgAuth
{
    MessageAuthentication        HMAC_SHA2_256
    MessageAuthentication        HMAC_SHA2_384
}
ConnectionDescriptorGroup connDescGrp
{
      ConnectionDescriptor
      {
            Protocol                TCP
            LocalPortRange          1024-65535
            RemotePortRange         21
            Jobname                 JOBNAME1
            TcpConnectionDirection  Inbound
      }
      ConnectionDescriptor
      {
            Protocol                TCP
            LocalPortRange          1024-65535
            RemotePortRange         50000-50200
            Jobname                 JOBNAME2
            TcpConnectionDirection  Outbound
      }
}
```

This rule is represented as a CNF rule with the following condition levels (levels are ANDed together):

- Level 1 = local address 9.1.1.1 OR local address 10.1.1.1
- Level 2 = remote address 200.1.1.1.1 OR remote address 201.1.1.1
- Level 3 = (local port range 1024-65535 AND remote port range 21 AND job name JOBNAME1 AND TCP connection direction Inbound) OR (local port range 1024-65535 AND remote port range 50000-50200 AND job name JOBNAME2 AND TCP connection direction Outbound)
- Level 4 = TLSv1.2 OR TLSv1.3
- Level 5 = SymmetricEncryption AES_CBC_128 OR SymmetricEncryption AES_CBC_256 OR SymmetricEncryption AES_GCM_256
- Level 6 = MessageAuthentication HMAC_SHA2_256 OR MessageAuthentication HMAC_SHA2_384

The **pasearch** command displays the ZERT policy as a complex rule. The security protection conditions (Level 4, 5, and 6 in this example) are only displayed in the summary condition.

Policy configuration files

A single file, the main configuration file, is specified explicitly or by default when the Policy Agent is started. This main configuration file points to other configuration files, as shown in Figure 109 on page 807.

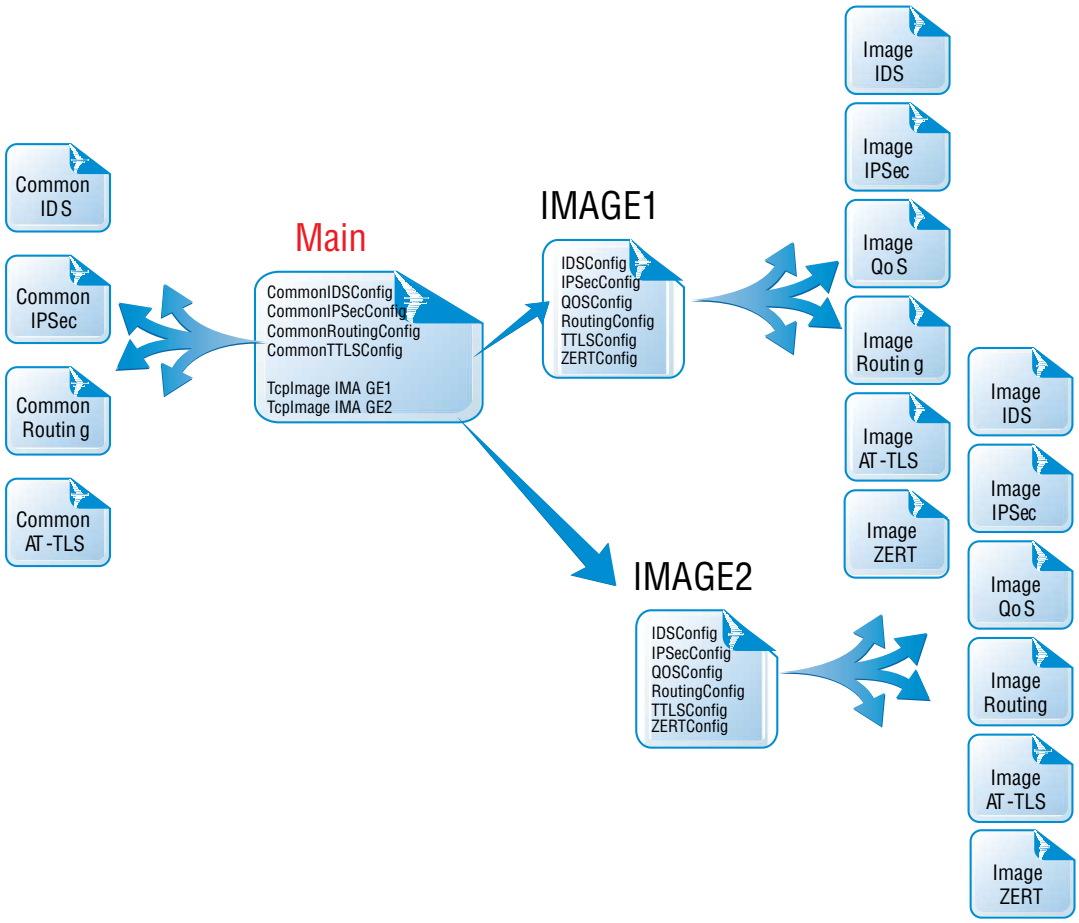


Figure 109. Policy Agent configuration files

For more information about the Policy Agent search order, see z/OS Communications Server: IP Configuration Reference.

You can specify statements in all configuration files using a variety of EBCDIC code pages. Use the Codepage configuration statement in the main configuration file to specify the code page to be used for all configuration files. The default code page is IBM-1047.

The main configuration file contains TcpImage or PEPInstance statements that define the TCP/IP stacks to be configured. The Policy Agent reads and installs policies for this set of stacks. Each TcpImage or PEPInstance statement can optionally specify the file name of an image-specific configuration file for that stack. If this file name is not specified, the main configuration file is also the image-specific configuration file for that stack. Thus, a single physical file can serve as two distinct logical files.

The main configuration file can also contain statements that specify the file names of one or more common configuration files for certain policy types. For example, the CommonIDSConfig statement points to a file containing IDS configuration statements that can be used for all configured stacks.

Each image-specific configuration file can contain statements that specify the file names of image-specific configuration files for different policy types. For example, the IPSecConfig statement points to a file containing IPSec configuration statements to be used for the stack that is represented by the image configuration file that contains the IPSecConfig statement. QoS policies can optionally be configured directly in the image configuration file, without using a QOSConfig statement.

The main configuration file on the policy server can also contain DynamicConfigPolicyLoad statements that determine the configuration files used when remote policy clients request policies. Each DynamicConfigPolicyLoad statement can serve a single policy client or a group of policy clients. Both common and policy client-specific configuration files can be specified for each policy type.

Steps for configuring the Policy Agent

Details for each general step are described in a corresponding subtopic.

Before you begin

You need to understand the hierarchy and relationships of the different configuration files. For more information, see “Policy configuration files” on page 806.

Procedure

Perform the following steps to configure the Policy Agent:

1. Configure general information.
2. Configure Policy Agent as a policy server.
3. Configure Policy Agent as a policy client.
4. Configure policies in Policy Agent configuration files.
5. Configure Policy Agent to use the LDAP server using the ReadFromDirectory statement.
6. Configure Policy Agent for import services.
7. Configure Policy Agent for automatic monitoring of applications.

Step 1: Configure general information

Before defining policies, you need to configure some basic operational characteristics of the Policy Agent.

Procedure

Follow these steps to configure these characteristics:

1. Set the TZ and LIBPATH environment variables.

   Use the TZ environment variable to specify the correct time zone. Use the LIBPATH environment variable so that the required dynamic link library (DLL) files can be located when you start the Policy

Agent. For information about how to specify these environment variables, see "Starting and stopping the Policy Agent" on page 823. You can also refer to comments in the sample start procedure that is shipped in SEZAINST(EZAPAGSP).

2. Specify the name of the main configuration file.

   You can specify the name of the main configuration file using the following methods:

   - -c start option
   - PAGENT_CONFIG_FILE environment variable
   - Default file name /etc/pagent.conf

   For information about the search order that Policy Agent uses to locate the main configuration file, and for examples of different ways to specify the main configuration file name, see "Starting and stopping the Policy Agent" on page 823. You can also refer to comments in the sample start procedure that is shipped in SEZAINST(EZAPAGSP).

3. Define the TcpImage or PEPInstance statements in the main Policy Agent configuration file.

   The PEPInstance statement is a synonym for TcpImage, and either can be used. PEPInstance refers to policy enforcement point (PEP), the component of a policy system that enforces policies, which for z/OS is the TCP/IP stack.

   Results:

   - The refresh interval used for the main configuration file will be the smallest of the values specified for the image-specific configuration files.
   - When the main configuration file is an MVS data set, it is reread at each refresh interval (which is the smallest of the individual stack refresh intervals), regardless of whether it has actually been changed or not. Because Policy Agent restarts all stack-related processing when the main configuration file is reread, this effectively makes the refresh interval for all stacks the same as this smallest configured interval.
   - The TcpImage or PEPInstance statement and all its parameters have no effect on policies defined for policy clients.

   To install a common set of policies to a set of stacks served by a single Policy Agent, do not specify image-specific configuration files for each image. In this case, there is only one configuration file (the main one) and the policy information contained in it is installed to all of the configured stacks. Different refresh intervals can also be configured for each image, but would probably be less useful in this case.

   In either case, it is possible that TCP/IP stacks configured to the Policy Agent are not started or even defined. The Policy Agent will fail when trying to connect to those stacks and log appropriate error messages.

   Rule: To dynamically add a TCP/IP stack to the Policy Agent configuration and have active policies automatically installed, in addition to adding the TcpImage statement to the configuration file, further action might be necessary as follows:

   - If the Policy Agent was started with the `-i` startup option, no further action is necessary. Active policies will be automatically installed to the stack when it becomes active.
   - If the Policy Agent was not started with the `-i` startup option, take one of the following actions:
     - Issue the MODIFY REFRESH or MODIFY UPDATE command after the stack becomes active. If you issue the MODIFY REFRESH or MODIFY UPDATE command before the stack becomes active, policies are not automatically installed.
     - Wait on the next update interval to check for configuration changes. If the stack is not active, policies are not automatically installed.

   The Policy Agent does not end when any (or all) stacks end. When the stacks are restarted, active policies are automatically reinstalled.

The TcpImage statement specifies a TCP/IP image and its associated configuration file to be installed to that image. The following example installs the policy control file `/tmp/TCPCS.policy` to the TCPCS TCP/IP image, after flushing the existing policy control data:

```
TcpImage TCPCS /tmp/TCPCS.policy FLUSH
```

For information about the FLUSH, NOFLUSH, PURGE, and NOPURGE parameters, see "FLUSH and PURGE considerations" on page 827.

4. Define the log file destination and the appropriate logging level.

   You can specify that Policy Agent log messages should be written to the syslog daemon or to a z/OS UNIX file. To indicate the syslog daemon, specify SYSLOGD in uppercase. To indicate a z/OS UNIX file, specify the file name. Specify the log file destination using the -l start option or the PAGENT_LOG_FILE environment variable, or you can use the default value /tmp/pagent.log.

   Tip: Specify SYSLOGD to take advantage of a centralized logging mechanism.

   Result: If Policy Agent cannot read the start options, then it does not have a log file destination. Policy Agent might fail to open a z/OS UNIX log file. In these situations, Policy Agent logs error messages to the syslog daemon and exits abnormally.

   Guidelines: If you run Policy Agent with a nonzero UID and you are using a z/OS UNIX log file, follow these guidelines:

   - Specify the file permissions as either 776 or 766.
   - Make sure that the syslog daemon is not configured to log to the same z/OS UNIX file. The syslog daemon runs with UID 0, so Policy Agent might not be able to access its log file if the syslog daemon creates the file before Policy Agent starts.

   The LogLevel statement is used to define the amount of information to be logged by the Policy Agent. The default is to log only event, error, console, and warning messages. This might be appropriate for a stable policy configuration, but more information might be required to understand policy processing or debug problems when first setting up policies or when making significant changes. Specify the LogLevel statement with the appropriate logging level in the main configuration file.

   Note: The maximum logging level (511) can produce a significant amount of output, especially with large LDAP configurations. This is not a concern if a z/OS UNIX log file is used, because Policy Agent uses a set of log files with a finite size in a round-robin configuration (the number and size of these files is controllable with the PAGENT_LOG_FILE_CONTROL environment variable). But when using the syslog daemon as the log file, the amount of log output produced should be taken into consideration.

5. Provide the following security authorizations.

   Because the Policy Agent can affect system operation significantly, the following security authorizations are required.

   - A user starting Policy Agent must be a superuser.
   - Security product authority (for example, RACF(R)) is required to start the Policy Agent. For sample commands needed to create the profile name and permit users to it, see the EZARACF sample in SEZAINST.

6. If Policy Agent's PAPI clients, including the **pasearch** command, are not defined as a superuser, to retrieve policies you must define security product authority in the SERVAUTH class for that client.

   The security product authority is always required in cases where the image name cannot be defined as a superuser. Remote policy clients are never defined as a superuser on the policy server, so security product authority is always required for them. These profiles can be defined by image name and policy type (ptype = QoS, IDS, IPSec, TTLS, Routing, or ZERT). Using a wildcard for profile names is allowed.

   ```
   EZB.PAGENT.sysname.image.ptype
   ```

   where:

   - *sysname* - System name defined in sysplex

- *image* - Tcp name, policy client name, or import request name for policy information that is being requested
- *ptype* - Type that is being requested:
  - QOS - Policy QoS
  - IDS - Policy IDS
  - IPSec - Policy IPSec
  - TTLS - Policy AT-TLS
  - Routing - Policy Routing
  - ZERT – Policy ZERT
  - CFGSERV - TCP/IP profile information

For details about the import request name, see "Import services" on page 796.

Tip: You can specify a wildcard on segments of the profile name.

Rules:

- When you use policy clients, the image portion of the profile name on the policy server must match or include the name of the policy clients. Configure each policy client name using the ClientName parameter on the PolicyServer statement, or use the default value for the ClientName parameter. For information about specifying the client name on the PolicyServer statement, see z/OS Communications Server: IP Configuration Reference.
- When you use import requesters, the image portion of the profile name on the policy server must match or include the import request name. If you use the IBM Configuration Assistant for z/OS Communications Server as the import requester, configure the import request name on the request panels for discovery import. For information about the import request name, see "Import services" on page 796.

Policy Agent checks all client requests to verify that the SERVAUTH class is active and that the profiles exist for the images and types in the request.

If a client's request is for multiple images or policy types, and permission is granted for only a subset of what is requested, Policy Agent returns only information for the subset for which permission is granted. However, if permission is denied for an entire request, including instances when only a single image or policy type is requested, an error is returned to the client indicating that permission is denied.

If the SERVAUTH class is not active or profiles are not active for the client's request (image, policy type), an error is returned to the client indicating that permission is denied.

See the EZARACF sample in SEZAINST for sample commands needed to create the profile name and permit users to it.

Step 2: Configure Policy Agent as a policy server

Define the port for clients to connect, and policy client matching statements to select the configuration files to be used for clients. Create user IDs on the server to match the user IDs of clients, and set up security.

Procedure

If you want to use the Policy Agent as a policy server, perform these steps:

1. Define the port to which policy clients will connect.

   If policy clients are to be used, the ClientConnection statement in the main configuration file provides the port that Policy Agent listens on for remote connections. You can use the default port (16310), but you must specify the ClientConnection statement to use Policy Agent as a policy server.

   Guideline: Reserve the port specified on the ClientConnection statement using the PORT statement in the TCP/IP profile.

Restriction: The port value cannot match the port value that is configured on the ServicesConnection statement.

2. Define a set of policy client matching statements that select the configuration files to be used for policy clients.

   When a policy client connects to the policy server, the DynamicConfigPolicyLoad statements in the main configuration file are evaluated to determine whether there is a match. The names are case sensitive with regard to matching. When a matching statement is found, the parameters identify both common and image-specific configuration files to be used for the policy client. If no matching statement is found, a default image-specific file is used. A matching statement (or default values) is bound to a policy client for the life of that client, until one of the following events occur:

   - The policy client disconnects from the policy server.
   - The connection between the policy server and policy client ends.
   - The associated DynamicConfigPolicyLoad statement is removed. In this case, the policy client is bound to a different DynamicConfigPolicyLoad statement (or to default values).

   You can use a regular expression for the policy client name on the DynamicConfigPolicyLoad statement to cause the statement to match multiple policy clients. For a description of supported regular expressions on the DynamicConfigPolicyLoad statement, see z/OS Communications Server: IP Configuration Reference.

   For example, the expression (.+)_(.+) matches any client name composed of one or more characters, followed by an underscore, followed by one or more characters. The default client names configured on the PolicyServer statement on the policy client would match this expression.

   You can use two different methods to substitute all or part of the client name in parts of the image-specific file name.

   - A single wildcard character (*) is replaced with the entire client name.
   - If you use a regular expression as the DynamicConfigPolicyLoad statement client name, you can use the symbolic replacement values $0 through $9 in the image-specific file name. The value $0 represents the entire portion of the client name that matched, while the values $1 through $9 represent portions of the client name that match corresponding parenthesized sub-expressions in the regular expression. Using the regular expression (.+)_(.+) as an example, and a policy client name of SYS123_TCPIP2, the values of the possible replacement variables are as follows:
     - The value * is replaced with SYS123_TCPIP2
     - The value $0 is replaced with SYS123_TCPIP2
     - The value $1 is replaced with SYS123
     - The value $2 is replaced with TCPIP2

     In this example, the value $0 is the same as the value *, but this does not hold true for all regular expressions. If you want to use the entire client name as a replacement value, specify the value *.

   The matching hierarchy used is as follows:

   a. Exact match between the policy client name and the DynamicConfigPolicyLoad statement.
   b. Regular expression match between the policy client name and the DynamicConfigPolicyLoad statement. The longest matching regular expression is chosen. If multiple statements match with the same length *clientname* parameter, the statement chosen is based on alphabetical order.
   c. No matching statement. A default file is used based on the policy type, as follows:

      Policy type
      Default file

      AT-TLS
      /etc/pagent_remote.ttls

      IDS
      /etc/pagent_remote.ids

IPSec
/etc/pagent_remote.ipsec

QoS
/etc/pagent_remote.qos

Routing
/etc/pagent_remote.routing

ZERT
/etc/pagent_remote.zert

The following example shows the DynamicConfigPolicyLoad statement matching by using a regular expression to simulate a simple wildcard, and the resulting configuration files that are used, using IPSec policies.

```
DynamicConfigPolicyLoad  Rem.*
{
  PolicyType             IPSec
  {
    CommonPolicyLoad     //'ETC.COMMON.IPSEC'
    PolicyLoad           //'ETC.IPSEC(*)'
  }
}

DynamicConfigPolicyLoad  Remote.*
{
  PolicyType             IPSec
  {
    PolicyLoad           /etc/*.ipsec
  }
}

DynamicConfigPolicyLoad  Remote5
{
  PolicyType             IPSec
  {
    CommonPolicyLoad     /user10/common_remote.ipsec
    PolicyLoad           /user10/pagent_remote5.ipsec
  }
}
```

The resulting configuration files used for a variety of policy clients are shown in Table 40 on page 813:

Table 40. Configuration files used for various policy clients

Policy client name	Matching statement	Common IPSec configuration file	Image IPSec configuration file
Remote1	Remote.*	None	/etc/Remote1.ipsec
Remote5	Remote5	/user10/common_remote.ipsec	/user10/pagent_remote5.ipsec
Rem42	Rem.*	//'ETC.COMMON.IPSEC'	//'ETC.IPSEC(REM42)'
remote5	Not applicable	None	/etc/pagent_remote.ipsec

The following example shows the DynamicConfigPolicyLoad statement matching by using a more complex regular expression, and the resulting configuration files that are used, using IDS policies. The regular expression matches two strings separated by an underscore character. Each string must begin with an uppercase alphabetic character and end with a numeric character.

```
DynamicConfigPolicyLoad  ^([A-Z].+[0-9]+)_([A-Z].+[0-9]+)$
{
  PolicyType             IDS
  {
    CommonPolicyLoad     //'ETC.COMMON.IDS'
    PolicyLoad           //'ETC.$1($2)'
  }
}
```

The resulting configuration files used for a variety of policy clients are shown in Table 41 on page 814:

Table 41. Configuration files used for various policy clients

Policy client name	Matching statement	Common IDS configuration file	Image IDS configuration file
SYS42_TCPIP2	^([A-Z].+[0-9]+)_([A-Z].+[0-9]+)$	//'ETC.COMMON.IDS'	//'ETC.SYS42(TCPIP2)'
Remote1_Image5	^([A-Z].+[0-9]+)_([A-Z].+[0-9]+)$	//'ETC.COMMON.IDS'	//'ETC.REMOTE1(IMAGE5)'
SYS123_TCPIP	Not applicable	None	/etc/pagent_remote.ids

Rule: The policy client names and DynamicConfigPolicyLoad statement names are case sensitive, but MVS data set names are not. Therefore, be careful when defining MVS data set configuration files that include a wildcard that is to be substituted with the policy client name. For example, the policy client names `client42` and `Client42`, if used as a substitution variable in an MVS data set name, would result in the same configuration file being used for both policy clients.

3. Configure one or more user IDs on the policy server system to match the user IDs of the policy clients.

   Each policy client uses a unique client name, but also must present valid credentials to the policy server. Valid credentials include a user ID and password or a user ID and PassTicket (if secure signon is enabled).

   Rule: The password defined for the user ID must match the password configured using the AuthBy Password parameter on the PolicyServer statement on the policy client.

   A SAF user ID representing a policy client must be defined to the security product. The user ID must be defined with an OMVS segment. When RACF is used as the security product, define the SAF user ID with the following command:

   ```
   ADDUSER client PASSWORD(password) DFLTGRP(OMVSGRP) OMVS(UID(x) HOME('/home/
   client'))
   ```

   Each policy client does not need to use a unique user ID, although that is a configuration option. The user ID is used for two purposes on the policy server:

   - User authentication when the policy client connects to the policy server
   - Access to SERVAUTH profiles to determine which policies the client can access

4. Permit Policy Agent to the BPX.DAEMON FACILITY class profile.

   For information about the use of the BPX.DAEMON profile, see "BPX.DAEMON FACILITY class profile" on page 37. If you decide to use this profile, permit the Policy Agent user ID to it. When RACF is used as the security product, permit the user ID with the following command:

   ```
   PERMIT BPX.DAEMON CLASS(FACILITY) ID(userid) ACCESS(READ)
   ```

5. If you want to use PassTicket security for policy clients, configure PTKTDATA class profiles.

   Policy clients can be configured to use either a password or a PassTicket on the PolicyServer statement, used when they connect to the policy server. For information about the secured signon function and PassTickets, see z/OS Security Server RACF Security Administrator's Guide. If you choose to use PassTickets, define the appropriate profiles in the PTKTDATA class to contain the application key used to generate and validate the PassTicket. When RACF is used as the security product, define the profiles with the following command:

   ```
   RDEFINE PTKTDATA profile SSIGNON(KEYMASKED(key)) UACC(UPDATE)
   ```

   The application name used by Policy Agent is PAGENT, so you need to define a profile with this name. The application key defined in the profiles must be the same on the policy client and policy server.

6. If you want to use secure connections between policy clients and the policy server, configure AT-TLS rules on the policy server to enable SSL connections from the policy clients.

   If policy clients use SSL connections, you must define AT-TLS rules on the policy server for communications between the policy client and server to be secured using AT-TLS. AT-TLS processing for a stack is enabled by specifying the TTLS parameter on the TCPCONFIG statement in the TCP/IP profile. Specific AT-TLS policy is configured in Policy Agent configuration files. For details about enabling AT-TLS and configuring AT-TLS policy, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

   Rules:

   - Define AT-TLS policy such that only cipher suites requiring TLS encryption are exchanged with policy clients. Failure to restrict the cipher suites to those requiring encryption might result in sensitive information flowing in the clear across an untrusted network.
   - Define AT-TLS policy for each stack through which policy server and policy client communication can flow.
   - If some policy clients use SSL and others do not use SSL, define AT-TLS policy to select only those policy clients that use SSL.

   Result: If you choose not to use SSL for your policy client to policy server connections, sensitive information flows in the clear on the connections. Sensitive information includes, but is not limited to, the following information:

   - The password that is sent from the policy client to the policy server for authentication (if you are using password credentials)
   - Policy information that is sent from the policy server to the policy client, such as:
     - Passwords
     - Certificate labels
     - IPSec keys for IKE tunnels that use pre-shared keys
     - IPSec keys for manual tunnels

   Requirement: The policy server acts as the server during an SSL handshake. To act in the server role of an SSL handshake, the policy server must have access to a private key and certificate verifying its ownership of that private key. For information about creating and managing keys and certificates for servers that use AT-TLS, see Appendix B, “TLS/SSL security,” on page 1329.

   An example of the AT-TLS policy statements used to enable AT-TLS for the policy server is as follows:

```
TTLSRule                      PolicyServerRule
{
   LocalPortRange                       16310
   JobName                              PAGENT
   Direction                            Inbound
   TTLSGroupActionRef                   PolicyServerGroup
   TTLSEnvironmentActionRef             PolicyServerConn
}

TTLSGroupAction
PolicyServerGroup
{
   TTLSEnabled                          On
}

TTLSEnvironmentAction                   PolicyServerConn
{
   TTLSKeyRingParms
   {
      Keyring                           PAGENT/keyring
   }
   TTLSCipherParmsRef                   RequireEncryption
   HandshakeRole
SERVER
}

TTLSCipherParms                    RequireEncryption
{
```

```
   V3CipherSuites              TLS_DHE_RSA_WITH_3DES_EDE_CBC_SHA
   V3CipherSuites              TLS_DHE_DSS_WITH_3DES_EDE_CBC_SHA
   V3CipherSuites              TLS_DH_RSA_WITH_3DES_EDE_CBC_SHA
   V3CipherSuites              TLS_DH_DSS_WITH_3DES_EDE_CBC_SHA
   V3CipherSuites              TLS_RSA_WITH_3DES_EDE_CBC_SHA
   V3CipherSuites              TLS_DHE_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites              TLS_DHE_DSS_WITH_AES_256_CBC_SHA
   V3CipherSuites              TLS_DH_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites              TLS_DH_DSS_WITH_AES_256_CBC_SHA
   V3CipherSuites              TLS_RSA_WITH_AES_256_CBC_SHA
   V3CipherSuites              TLS_DHE_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites              TLS_DHE_DSS_WITH_AES_128_CBC_SHA
   V3CipherSuites              TLS_DH_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites              TLS_DH_DSS_WITH_AES_128_CBC_SHA
   V3CipherSuites              TLS_RSA_WITH_AES_128_CBC_SHA
   V3CipherSuites              TLS_DHE_RSA_WITH_DES_CBC_SHA
   V3CipherSuites              TLS_DHE_DSS_WITH_DES_CBC_SHA
   V3CipherSuites              TLS_DH_RSA_WITH_DES_CBC_SHA
   V3CipherSuites              TLS_DH_DSS_WITH_DES_CBC_SHA
   V3CipherSuites              TLS_RSA_WITH_DES_CBC_SHA
   V3CipherSuites              TLS_RSA_WITH_RC4_128_SHA
   V3CipherSuites              TLS_RSA_WITH_RC4_128_MD5
   V3CipherSuites              TLS_RSA_EXPORT_WITH_RC2_CBC_40_MD5
   V3CipherSuites              TLS_RSA_EXPORT_WITH_RC4_40_MD5
}
```

Rule: The LocalPortRange value on the TTLSRule statement must include the value specified on the ClientConnection statement in the policy server main configuration file.

7. If you use secure connections between any policy clients and the policy server, permit Policy Agent to the EZB.INITSTACK.*sysname*.*tcpname* SERVAUTH class profile.

   Because AT-TLS policies are used to enable SSL connections from policy clients, Policy Agent must be permitted to the EZB.INITSTACK.*sysname*.*tcpname* SERVAUTH class profile if any policy clients use SSL. For more details, see "TCP/IP stack initialization access control" on page 1122. When RACF is used as the security product, permit Policy Agent to the profile with the following command:

```
PERMIT EZB.INITSTACK.sysname.tcpname CLASS(SERVAUTH) ID(userid) ACCESS(READ)
```

Step 3: Configure Policy Agent as a policy client

You can use the Policy Agent as a policy client.

Procedure

If you want to use the Policy Agent as a policy client, perform the following steps:

1. Define the parameters needed to connect to the policy server using the ServerConnection statement in the main configuration file:
   - Specify the host name (or IP address) and port of a primary and an optional backup server.
   - If you want to use a secure connection to the policy server, specify parameters for a secure SSL connection. For details, see "Add TLS/SSL to Policy Agent connections" on page 822.

   Requirement: Connectivity to the policy server is required for all images on the policy client that need to connect to the policy server.
2. Define the policy server parameters to be used for each image on the PolicyServer statement in the image configuration files:
   - When the policy client connects to the policy server, the policy client needs to supply a user ID and authentication information (password or PassTicket). Specify these parameters on the PolicyServer statement. The user ID must be defined on the policy server system. For information about the PTKTDATA class profiles that are needed when a PassTicket is specified on the PolicyServer statement, see step "5" on page 814 in "Step 2: Configure Policy Agent as a policy server" on page 811.
   - The policy server determines what configuration files to load based on a matching DynamicConfigPolicyLoad statement in its configuration. Specify the client name that the policy server is to use for matching. If this parameter is not specified, the default value is the policy client's

system name concatenated to the image name with an intervening underscore character (_). For example, if the client's system name is SYS42 and the image name for this policy client is TCPIP2, the default client name presented to the policy server is SYS42_TCPIP2.

- Specify the types of policies to be retrieved from the policy server. You can specify one or more policy types. You can also specify parameters for each policy type (FLUSH, NOFLUSH, PURGE, or NOPURGE). These parameters have the same meaning as the corresponding parameters on the TcpImage or PEPInstance statement.

  For each policy type specified, the corresponding *xxx*Config statement for that type is ignored in the local configuration. For example, if PolicyType IPSec is specified on the PolicyServer statement, the IPSecConfig statement is ignored. This is true even if the primary and backup policy servers cannot be reached. You can use local or remote policy for each policy type, but not both.

Step 4: Configure policies in Policy Agent configuration files

Policies can be defined in any referenced Policy Agent configuration file. For more information, see the appropriate information for each policy type:

- Policy-based routing (See "Policy-based routing" on page 349)
- Quality of Service (See Chapter 15, "Quality of service," on page 831)
- Intrusion detection services (See Chapter 16, "Intrusion detection services," on page 851)
- IP filtering, and manual and dynamic VPN tunnels, collectively referred to as IPSec policies (See Chapter 17, "IP security," on page 883)
- Application Transparent Transport Layer Security (See Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121)
- zERT policy-based Enforcement (See "Defining a zERT enforcement policy" on page 176)

Step 5: Configure Policy Agent to use the LDAP server using the ReadFromDirectory statement

The ReadFromDirectory statement in the Policy Agent configuration file initializes the Policy Agent as an LDAP client. The policies are downloaded from the LDAP server, along with the policies specified in the Policy Agent configuration files.

When configuring the ReadFromDirectory statement, first specify the name (or IPv4 address) and port of the primary server and the same for the backup server (if one is used).

Notes:

1. The LDAP client library used to connect to the LDAP server does not support IPv6.
2. When using the z/OS LDAP server, the server listens on a separate port for SSL connections. This means that you should specify the correct port depending on whether or not SSL is used.

Next, configure other connection attributes. The Policy Agent (as an LDAP client) must log in to the LDAP server. The user ID and password for logging in must be configured on the ReadFromDirectory statement. The user ID is also known as Distinguished Name for user ID, and it is in the form of an LDAP DN. If the user ID and password are not specified, the Policy Agent uses anonymous login to connect to the server.

The LDAP server can be configured with only LDAP protocol version 3. To use LDAP protocol version 3, you can set LDAP_ProtocolVersion to 3 on the ReadFromDirectory statement. This is the default value. This statement also configures the version of the schema to be retrieved from the server.

Finally, configure attributes to indicate how to search the LDAP server for policies. Policy roles allow one or more roles, or role-combinations, to be assigned to policy rules using the ibm-policyRoles attribute. These roles represent the intended usage of the policy rules. For example, a role of "East Coast WAN" might be used to represent policies for the wide area network on the US East coast for an enterprise. Policy role values are not standardized; they are simply values used to assign roles to policies. When an entity that requires policies (such as Policy Agent) requests policies from an LDAP server, it can filter out policy rules that do not match the roles that it plays. Although similar to policy keywords, which also allow

search scoping, policy roles are a bit more sophisticated. Specifically, role-combinations are allowed, which take the form of a specification like "roleA && roleB", meaning both roleA AND roleB. Because the ibm-policyRoles attribute is multi-valued, a form of CNF/DNF logic can be used for policy roles: the roles in a role-combination are ANDed, and the roles or role-combinations specified on different values of this attribute are ORed.

For the Version 1 schema, a base DN to start the search, and a *selector tag* value are configured. The selector tag is used to match against the SelectorTag attribute in the policy objects. For Version 1, the Policy Agent also automatically includes the stack name when searching for policies; this value is matched against the TcpImageName attribute in the policy objects. For the Version 2 schema, a base DN to start searching is also configured. This DN can specify a single LDAP object, a policy group, or an LDAP subtree containing many objects. For filtering the search, three keywords can be configured:

- SearchPolicyKeyword matches against the ibm-policyKeywords attribute in any policy object.
- SearchPolicyGroupKeyword matches against the ibm-policyGroupKeywords attribute in policy group objects.
- SearchPolicyRuleKeyword matches against the ibm-policyRuleKeywords attribute in policy rule objects.

Optionally, specify parameters for a secure SSL connection. For details, see "Add TLS/SSL to Policy Agent connections" on page 822.

The example that follows this list takes the following actions:

- Connects to the LDAP server at IP address 9.100.1.1, using the default port 389.
- Specifies a user ID and password to log in to the server.
- Specifies LDAP protocol version 3.
- Specifies schema version 3.
- Starts searching at the DN ou=policy, o=IBM, c=US object/subtree.
- Selects only policy objects that contain either the "POLICY" or "EASTERN" keywords.

```
ReadFromDirectory
{
LDAP_Server 9.100.1.1
LDAP_DistinguishedName cn=root, o=IBM, c=US
LDAP_Password 4qr56jb
LDAP_ProtocolVersion 3
LDAP_SchemaVersion 3
SearchPolicyBaseDN ou=policy, o=IBM, c=US
SearchPolicyKeyword POLICY
SearchPolicyKeyword EASTERN
}
```

Step 6: Configure Policy Agent for import services

You can connect an import requester to the Policy Agent to provide TCP/IP information import services.

Procedure

To configure the Policy Agent for import services, perform the following steps:

1. Define the port and TCP/IP image name to which import requesters will connect.

   If import requesters are to be used, the ServicesConnection statement in the main configuration file provides the port and TCP/IP image name that the Policy Agent listens on for remote connections. An import requester is one type of services requester provided for by the ServicesConnection statement. The Policy Agent listens for services requester connections on only one TCP/IP image. You can specify the image name to be used, or use the name specified (or specified by default) on the TCPIPUSERID statement or TCPIPJOBNAME statement in TCPIP.DATA. If the default TCP/IP image cannot be determined, the Policy Agent uses the image name INET. In any case, the image name might or might not match an image name specified on a TcpImage statement:

- If the specified name does not match any TcpImage statement, the Policy Agent generates an internal TcpImage statement with default values to represent the TCP/IP image. This means that you can specify a maximum of only 7 (instead of 8) TcpImage or PEPInstance statements.
- In a single stack (INET) environment, the Policy Agent uses the active TCP/IP image to listen for services connection requests.

Rule: The ServicesConnection statement is required for any Policy Agent that accepts connections from an import requester.

Guideline: Reserve the port specified on the ServicesConnection statement using the PORT statement in the TCP/IP profile.

Restriction: The port value cannot match the port value configured on the ClientConnection statement.

2. Optionally configure secure connections from the import requesters.
   - By default, the ServicesConnection statement defines a basic connection that is not explicitly secured. This option can be used for an unsecure connection, or you can define AT-TLS policies for this import services connection to create a secure SSL connection.
   - You can define a secure connection instead, and specify the level of tracing and the TLS/SSL key ring to use. You must specify the name of a SAF key ring. Key ring files created by the System SSL gskkyman utility are not supported. When you configure a secure connection, Policy Agent automatically creates an AT-TLS policy for the connection, and the import requester must also specify that the connection is to be secured. You must enable the TTLS parameter on the TCPCONFIG statement in the TCP/IP profile for the generated AT-TLS policy to be effective.

Tip: This option only supports TLSv1.0 and TLSv1.1. It is not recommended for secure TLS/SSL. It is recommended for secure TLS/SSL to configure Security Basic and to supply user defined AT-TLS policies.

The following example shows a ServicesConnection statement for a secure connection:

```
ServicesConnection
{
   Port       17000
   ImageName  TCPIP1
   Security   Secure
   Trace      14
   Keyring    PAGENT/PAGRING
}
```

The following AT-TLS policy is generated from this ServicesConnection statement:

```
TTLSRule                       TTLS_RULE_____________GENERATED
{
   LocalPortRange              17000
   JobName                     PAGENT
   Direction                   Inbound
   TTLSGroupActionRef          TTLS_GROUP_ACTION_______GENERATED
   TTLSEnvironmentActionRef    TTLS_ENVIRONMENT_ACTION_GENERATED
}


TTLSGroupAction                TTLS_GROUP_ACTION_______GENERATED
{
   TTLSEnabled                 On
   Trace                       14
}

TTLSEnvironmentAction          TTLS_ENVIRONMENT_ACTION_GENERATED
{
   HandshakeRole               SERVER
   TTLSKeyRingParms
   {
      Keyring                  PAGENT/PAGRING
   }
}
```

The Policy Agent installs this generated policy in the TCP/IP image specified explicitly or by default on the ServicesConnection statement. This generated policy uses a priority value that is lower than any specified AT-TLS policies, so it is installed as the last policy in the TCP/IP image. If local or remote AT-TLS policies are configured, the Policy Agent installs those policies before installing the generated policy. If you configure AT-TLS policies on a policy server, those policies must be successfully retrieved before the Policy Agent is able to accept connections from services requesters. Accepting connections from services requesters can be affected by problems or delays in retrieving the AT-TLS policies from the policy server.

If you change the ServicesConnection statement, the generated policy is uninstalled or reinstalled as follows:

- If you change the Port, Trace or Keyring parameters, the Policy Agent regenerates and reinstalls the policy.
- If you change the ImageName parameter, the Policy Agent uninstalls the generated policy from the previous image and installs the policy on the new image.
- If you change the Security parameter value from Secure to Basic, the Policy Agent uninstalls the generated policy.

If you delete the ServicesConnection statement, the Policy Agent uninstalls the generated policy.

3. To restart the listen for services requester connections and, if required, to reinstall the generated AT-TLS policy, issue the MODIFY SRVLSTN command. For information about when you might use the MODIFY command for Policy Agent, see z/OS Communications Server: IP System Administrator's Commands.

Step 7: Configuring Policy Agent to automatically monitor applications

You can use the Policy Agent to automatically start, stop, and monitor a set of related applications. Policy Agent starts the applications and monitors them to ensure that they remain active.

Before you begin

If Policy Agent determines that any applications have not started or have stopped, it continues to try to start or restart the applications, up to a configurable retry limit within a configurable retry period.

Requirement: To automatically monitor applications, you must start Policy Agent with a user ID that has superuser authority UID(0). For sample RACF commands, see the EZARACF member of SEZAINST.

Procedure

To configure the Policy Agent for automatic monitoring, perform the following steps:

1. Decide what applications you want to monitor.

   You can use the Policy Agent to monitor any or all of the following applications:

   - Defense Manager daemon (DMD)
   - Internet Key Exchange daemon (IKED)
   - Network security services daemon (NSSD)
   - Syslog daemon (SYSLOGD)
   - Traffic Regulation Management daemon (TRMD)

   Determine which of these applications you currently use, or want to start using, in your environment, and for each application, determine whether you want the Policy Agent to start, stop, and monitor the application.

   Requirements:

   - To start the application, you must use a cataloged procedure that accepts a number of variables that are provided by the Policy Agent. A sample procedure is included in SEZAINST(EZAPOLPR).

- If you want to manually start, restart, or stop the application, you must use MODIFY commands that are directed to the Policy Agent. If you issue the commands directly to the application itself, Policy Agent is not aware of the action and the monitoring logic will probably not produce the expected results.

Results:

- If you start the Policy Agent after you have already started an application to be monitored, Policy Agent starts monitoring the application if it was originally started with the same job name that is configured to the Policy Agent. If the application needs to be restarted later, it is restarted using the cataloged procedure configured to the Policy Agent. This might not be the same procedure that was originally used to start the application.
- If you start the Policy Agent after you have already started an application to be monitored, but the application does not use the same job name that is configured to the Policy Agent, then the Policy Agent is not able to detect that the application is active. The Policy Agent will try to start another instance of the application, which is likely to fail.

Tip: If you configure applications to be monitored by the Policy Agent, ensure those applications are not running before starting the Policy Agent. However, you probably want to start syslogd before starting the Policy Agent, so you should ensure that Policy Agent is configured with the correct syslogd job name.

2. Configure the applications that you want to monitor using the AutoMonitorApps statement.

   You can configure applications that you want to monitor that are or are not associated with a particular TCP/IP stack. You can specify the cataloged procedure used to start each application, the job name for the application, and other application-specific parameters on the AutoMonitorApps statement.

   Perform the following steps to configure the applications that you want to monitor:

   a. Specify the AutoMonitorApps statement in the main Policy Agent configuration file.
      - Use the AppName parameter to specify each application that is not associated with a particular TCP/IP stack. All supported applications except TRMD fall into this category.
      - Use the TcpImageName and AppName parameters to specify each application that is associated with a particular TCP/IP stack. TRMD is the only application that falls into this category.
   b. Use the ProcName parameter for each AppName parameter on the AutoMonitorApps statement to specify the cataloged procedure that is used to start each application. Because all key data is passed to the procedure as variables, you can use a single procedure for all configured applications. You can also use a unique procedure for one or more applications.
   c. Use the Jobname parameter for each AppName parameter on the AutoMonitorApps statement to specify the job name for each application.
   d. Use the StartParms parameter for each AppName parameter on the AutoMonitorApps statement to specify start parameters for each application.
   e. Use one or more EnvVar parameters for each AppName parameter on the AutoMonitorApps statement to specify application-specific parameters, such as time zone or configuration file name. You can specify any or all environment variables that are accepted by the specific application.

   The following example shows the AutoMonitorApps statement:

```
AutoMonitorApps
{
  AppName            IKED
  {
    Procname         POLPROC
  }
  AppName            TRMD
  {
    TcpImageName     TCPIP1
    {
      Procname       POLPROC
      Jobname        TRMD1
    }
    TcpImageName     TCPIP3
    {
```

```
      Procname      POLPROC
      Jobname       TRMD3
    }
  }
}
```

This example shows how to specify parameters for two types of applications:

- An application without stack affinity, meaning that a single copy of the application runs regardless of how many TCP/IP stacks are running. This example uses IKED as such an application.
- An application with stack affinity, meaning that one instance of the application runs on each TCP/IP stack. This example uses TRMD as such an application.

3. Configure global monitoring parameters using the AutoMonitorParms statement.

   Use the AutoMonitorParms statement in the main Policy Agent configuration file to specify global monitoring parameters, such as the monitor time interval and retry limits.

   - Use the MonitorInterval parameter to specify the monitor interval in seconds.
   - Use the RetryLimitCount and RetryLimitPeriod parameters to specify how many times within a given time period Policy Agent should try to start or restart an application. If the application fails to successfully start or restart after the retry limit has been reached, Policy Agent stops trying until the application is manually started using the MODIFY *procname*,MON,START,*application* command.

Add TLS/SSL to Policy Agent connections

The Transport Layer Security (TLS) protocol is defined by the Internet Engineering Task Force (IETF) in RFCs 2246, 4346, 5246, and 8446. TLS is based on the Secure Sockets Layer (SSL) protocol. The TLS/SSL protocol begins with a handshake. During the handshake, the client authenticates the server, the server optionally authenticates the client, and the client and server agree on how to encrypt and decrypt information.

Server Authentication: When using TLS/SSL to secure communications, the TLS/SSL authentication mechanism known as server authentication is used. With server authentication, the server must have a digital certificate that authenticates the server to the Policy Agent client. The server supplies the client with the certificate during the initial TLS/SSL handshake. If the client validates the server's certificate, a secure communication channel is established between the server and the Policy Agent client.

For server authentication to work, the server must have a private key and associated server certificate in the server key ring file.

To conduct commercial business on the Internet, you might use a widely known Certificate Authority (CA), such as VeriSign, to get a high assurance certificate. For a relatively small private network within your own enterprise or group, you can issue your own certificates, called self-signed certificates, for your own use.

Client Authentication: When using TLS/SSL Client Authentication, the client passes a digital certificate to the server as part of the TLS/SSL handshake. To pass authentication, the Certificate Authority (CA) that signed the client certificate must be considered trusted by the server.

Self-signed Server Certificates: Normally, a server certificate should be obtained from a known CA. However, for testing, an installation might use a self-signed server certificate. Because the clients will not know about the issuer of the self-signed server certificate, in most cases it is necessary to add the server's self-signed certificate to the client's signer certificates.

The gskkyman utility can be used to create public/private key pairs and certificate requests, receive certificate requests into a key ring, and manage keys in a key ring. The gskkyman utility is documented in z/OS Cryptographic Services System SSL Programming. The gskkyman utility is shipped with z/OS in System SSL, which is part of the cryptographic services base element of z/OS. For detailed instructions on setting up certificates and key rings, see Appendix B, “TLS/SSL security,” on page 1329.

The set of SSL protocol cipher specifications to be allowed for the secure session can be set for the policy server connection.

Rule: Define AT-TLS policy on the policy server such that only cipher suites requiring TLS encryption are exchanged with policy clients. Failure to restrict the cipher suites to those requiring encryption might result in sensitive information flowing in the clear across an untrusted network.

For the list of cipher suites supported and the default order used if none is specified, see z/OS Cryptographic Services System SSL Programming.

The Policy Agent connection to LDAP can be secured using SSL by tailoring the following parameters on the ReadFromDirectory statement. This allows for protection of policy retrieval from an LDAP server.

- LDAP_SSLKeyringFile
- LDAP_SSLKeyringPassword
- LDAP_SSLName

The policy client connection to the policy server can be secured using TLS/SSL by tailoring the following parameters on the ServerConnection statement. This allows for protection of policy retrieval from the policy server.

- ServerSSLKeyring
- ServerSSLKeyringPassword
- ServerSSLKeyringStashFile
- ServerSSLName
- ServerSSLV3CipherSuites
- ServerSSLV3CipherSuites4Char
- ServerTLSv1.3

You can secure the connection used by services requesters by creating AT-TLS policy rules for the services requestor and services provider. Configure Security Basic on the ServicesConnection statement.

For more detail about these parameters, see z/OS Communications Server: IP Configuration Reference. Additional information about the concepts of cryptography and SSL can be found at the following websites:

- http://httpd.apache.org/docs/current/ssl/ssl_intro.html
- http://www.verisign.com/ssl/ssl-information-center/how-ssl-security-works/index.html

Starting and stopping the Policy Agent

You can start the Policy Agent in any of the following ways:

- From the z/OS shell

  If you use the shell, start the Policy Agent in a background shell session by specifying a trailing & on the command line.
- As a started task
- Using the COMMND*xx* member of parmlib

  Using the COMMND*xx* member of parmlib enables Policy Agent to be automatically started when the system is IPLed. For more information about COMMND*xx*, see z/OS MVS Initialization and Tuning Reference.
- In some cases, using the TCP/IP AUTOLOG statement

  For more information, see “AUTOLOG considerations” on page 823.

AUTOLOG considerations

If a procedure in the AUTOLOG list also has a PORT statement reserving a TCP or UDP port but does not have a listening connection on that port, TCP/IP periodically attempts to cancel that procedure and start it again.

Guideline: Do not use AUTOLOG to start the Policy Agent if any listening connections are used. The Policy Agent is unlike typical servers, and performs functions outside the realm of listening for connections from other applications. Sometimes during normal operation of the Policy Agent, listening connections are unavailable for short periods of time. When listening connections are unavailable, it is possible that the AUTOLOG timer could restart the Policy Agent.

Policy Agent optionally listens on one or more of the following ports:

- The pagentQosListener port (usually defined in the /etc/services file as port 1700). This happens only when the PolicyPerfMonitorForSDR statement is configured for a given TCP/IP stack.
- The port defined or specified by default on the ClientConnection statement, which is used to listen for remote policy client connections on all TCP/IP stacks.
- The port defined or specified by default on the ServicesConnection statement, which is used to listen for services requester connections on a single TCP/IP stack. For information about the ServicesConnection statement and specifying which stack to use to listen for these connections, see z/OS Communications Server: IP Configuration Reference.

If you want to start Policy Agent with AUTOLOG, you must take one of the following actions:

- Ensure that none of the listening ports used by Policy Agent are reserved by the PORT statement in the TCP/IP profile.
- Add the NOAUTOLOG parameter to the PORT statement in the TCP/IP profile. For example:

```
PORT
 1700  TCP PAGENT NOAUTOLOG
 16310 TCP PAGENT NOAUTOLOG
 16311 TCP PAGENT NOAUTOLOG
```

In addition, when the PolicyPerfMonitorForSDR statement is being used, if the pagentQosCollector port (usually defined in the /etc/services file as port 1701) is reserved in the PORT list it should always be specified with the NOAUTOLOG parameter, because this port is never used as a listening port. For example:

```
PORT
 1701 TCP PAGENT NOAUTOLOG
```

Tip: When Policy Agent is not listening on any ports, you can use the autostart feature of AUTOLOG as previously described. However, the monitoring and automatic restart features of AUTOLOG are unavailable because AUTOLOG must listen to a TCP or UDP connection.

If you fail to take one of the above actions, Policy Agent will be periodically canceled and restarted by TCP/IP.

Specifying environment variables

The Policy Agent requires access to one or more DLLs at run time. The LIBPATH environment variable needs to be set to include the /usr/lib directory, which normally includes all the required DLLs.

For policy time specifications to be properly acted upon, the TZ environment variable needs to be set to local time. You can set the LIBPATH and TZ environment variables as follows:

- When starting from the z/OS shell:

  Export the LIBPATH and TZ environment variables before starting the Policy Agent. This is best accomplished in /etc/profile or in .profile in the HOME directory. For example, if you are in the Eastern time zone in the United States:

  ```
  export LIBPATH=/usr/lib
  export TZ=EST5EDT
  ```

- When starting as a started task, use either of the following methods:
  - Specify LIBPATH and TZ using the ENVAR parameter on the PARM statement in the started procedure. For example:

```
// PARM=('ENVAR("LIBPATH=/usr/lib","TZ=EST5EDT")/')
```

– Export the LIBPATH and TZ environment variables in a file specified with the STDENV DD statement. For example:

```
// PARM='ENVAR("_CEE_ENVFILE=DD:STDENV")/'
//STDENV DD PATH='/etc/pagent.env',PATHOPTS=(ORDONLY)
```

In the /etc/pagent.env file:

```
LIBPATH=/usr/lib
TZ=EST5EDT
```

The use of the STDENV DD statement works well when you want to specify more than one environment variable; there is a JCL limit of 100 characters on the PARM parameter. Language Environment recommends a variable record format for the STDENV file.

You can also set the TZ environment variable for all applications in the CEEPRM*xx* PARMLIB member. You should define the TZ environment variable for all three LE option sets (CEEDOPT, CEECOPT, and CELQDOPT). For example:

```
CEECOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
CEEDOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
CELQDOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
```

For more information on specifying run-time options, see z/OS Language Environment Programming Guide. For details on setting the LIBPATH and TZ environment variables, see z/OS UNIX System Services Command Reference.

Main configuration file search order

Although the /etc/pagent.conf is the default configuration file, a specific search order is used when starting the Policy Agent. The following order is used:

1. File or data set specified with the `-c` startup option
2. File or data set specified with the PAGENT_CONFIG_FILE environment variable
3. /etc/pagent.conf

The syntax for a z/OS UNIX file is different than the syntax for an MVS data set. The following examples use the PAGENT_CONFIG_FILE environment variable:

- PAGENT_CONFIG_FILE=/*dir*/*file*
- PAGENT_CONFIG_FILE=//'*mvs.dataset.name*'

Other considerations when starting the Policy Agent

If you start the Policy Agent with a user ID that does not have superuser authority [UID(0)], then read and write permission is required for the following directories and files:

- Directories
  - /tmp/
  - /var/tmp/
- Files
  - Files created when you configure the PerformanceLogFile parameter on the PolicyPerformanceCollection statement

    For information about the PolicyPerformanceCollection statement, see z/OS Communications Server: IP Configuration Reference.
  - Policy Agent log files

For information about using the -l parameter when starting Policy Agent from the z/OS shell to specify the destination for the log file, see z/OS Communications Server: IP Configuration Reference.

– Policy Agent pid file

The /tmp/pagent.pid is a temporary file that the Policy Agent creates. This file contains the process ID of the current invocation of the Policy Agent.

Restrictions:

- If /tmp/pagent.pid is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the Policy Agent.
- If /tmp/pagent.pid is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /tmp/pagent.pid is stored cannot have write access to the directory. Additionally, write access to /tmp/pagent.pid must be limited to the owning UID or group, for example, --w--w----permissions.

Requirement: To automatically monitor applications, you must start Policy Agent with a user ID that has superuser authority UID(0). For sample RACF commands, see the EZARACF member of SEZAINST.

At initialization, the Policy Agent creates a z/OS UNIX file called /tmp/*tcpname*.Pagent.tmp. This occurs for every TCP/IP stack defined on a TcpImage statement.

In this z/OS UNIX file, *tcpname* is the name of a TCP/IP stack from a TcpImage statement. During TCP/IP stack initialization, the TCP/IP stack will attempt to modify a file by this name to notify the Policy Agent that the stack has been reactivated. This causes the Policy Agent to automatically attempt to reinstall the existing policies to this stack.

When /tmp/*tcpname*.Pagent.tmp is in a read/write sysplex-aware z/OS file system (zFS), a symbolic link can be created to a file that is in a read-only zFS to enable the update notification. In the following example, the **ln** command is used to create a symbolic link to a file in /mydir, where /mydir is in a read-only zFS:

```
ln -s /mydir/TCPIP.Pagent.link /tmp/TCPIP.Pagent.tmp
```

Restrictions:

- Policy Agent
  - If /tmp/*tcpname*.Pagent.tmp is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the Policy Agent.
  - If /tmp/*tcpname*.Pagent.tmp is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /tmp/*tcpname*.Pagent.tmp is stored cannot have write access to the directory. Additionally, write access to /tmp/*tcpname*.Pagent.tmp must be limited to the owning UID or group, for example, --w--w----permissions.
- TCP/IP stack
  - If /tmp/*tcpname*.Pagent.tmp is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the TCP/IP stack.
  - If /tmp/*tcpname*.Pagent.tmp is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /tmp/*tcpname*.Pagent.tmp is stored cannot have write access to the directory.

To ensure that only one Policy Agent is started, the Policy Agent uses the following enqueue:

- Enqueue name is TCP_TCPI
- Resource name is TCPIP.PAGENT

When starting from the shell, note that the Policy Agent executable file is in the /usr/lpp/tcpip/sbin directory. There is also a link from the /usr/sbin directory. Make sure your PATH statement contains either the /usr/sbin or /usr/lpp/tcpip/sbin directory.

For example, the following command starts Policy Agent with these characteristics:

```
pagent -c /u/user10/pldap.conf -l SYSLOGD &
```

- Policy Agent uses the configuration file /u/user10/pldap.conf
- Policy Agent logs output to the syslog daemon (SYSLOGD). Note that "SYSLOGD" must be specified in uppercase to obtain this behavior

Use the S PAGENT command on an MVS console or SDSF to start the Policy Agent as a started task. A sample procedure is shipped in member EZAPAGSP in SEZAINST.

Stopping the Policy Agent

You can stop the Policy Agent by:

- Using the operator command STOP
- Using the kill command in the z/OS shell
- Using the operator command CANCEL. Use the CANCEL command only as a last resort if the STOP or kill commands do not completely stop the Policy Agent.

Result: When the Policy Agent is shut down normally (KILL or STOP), if the PURGE option is configured, all QoS, IDS, and AT-TLS policies are purged from this stack. IPSec, Routing, and ZERT policies are not automatically purged.

The following kill command with the TERM signal will enable Policy Agent to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

where pid is the Policy Agent process ID.

The Policy Agent process ID can be obtained using the following z/OS UNIX command:

```
 ps -A
```

It can also be obtained from the /tmp/pagent.pid file. The /tmp/pagent.pid file is a temporary file created by the Policy Agent. It contains the process ID of the current invocation of the Policy Agent. This temporary file is deleted when the Policy Agent is stopped.

Refreshing policies

The MODIFY command may be used to interactively cause the Policy Agent to reread the configuration information and, if requested, download objects from the LDAP server. In addition to this, the Policy Agent will also accept SIGHUP signals to perform the refresh function. See the z/OS Communications Server: IP System Administrator's Commands for more detailed information on the MODIFY command.

FLUSH and PURGE considerations

The FLUSH/NOFLUSH and PURGE/NOPURGE parameters can be configured for each policy type supported by the Policy Agent.

These parameters determine whether or not policies are deleted from the associated TCP/IP stack under certain conditions, as detailed in Table 43 on page 828.

Table 42 on page 827 shows where you configure these parameters for each type of local or remote policy.

Table 42. Where Policy Agent FLUSH and PURGE are configured

Policy type	**Statement where configured**
Local Routing and ZERT policies	Not configurable (always support FLUSH and NOPURGE)

Table 42. Where Policy Agent FLUSH and PURGE are configured (continued)	
Policy type	**Statement where configured**
Local IDS policies	IDSConfig or TcpImage/PEPInstance
Local IPSec policies	Not supported
Local QoS policies	TcpImage/PEPInstance
Local AT-TLS policies	TTLSConfig or TcpImage/PEPInstance
Remote policies (all types except IPSec, Routing, and ZERT)	PolicyServer or TcpImage/PEPInstance

Results:

- IPSec policies do not use these parameters. Instead, IPSec functions as though the FLUSH and NOPURGE parameters are always specified, with the exception that the FLUSH parameter has no effect when the MODIFY REFRESH command is entered.
- Parameters specified on the TcpImage/PEPInstance statement are overridden by parameters configured on other statements.

Table 43 on page 828 shows the results of using the FLUSH and PURGE parameters.

Table 43. How Policy Agent FLUSH and PURGE are used			
Event	**IPSec policies**	**Routing and ZERT policies**	**Other policies**
Policy Agent start (FLUSH defined)	All policies are replaced in the TCP/IP stack.	All policies are deleted and reloaded into the TCP/IP stack.	All policies are deleted and reloaded into the TCP/IP stack.
Policy Agent start (NOFLUSH defined)	All policies are replaced in the TCP/IP stack.	All policies are deleted and reloaded into the TCP/IP stack.	All changed policies are updated in the TCP/IP stack. Deleted policies are not removed from the TCP/IP stack.
Policy Agent termination (PURGE defined)	TCP/IP stack policies are unchanged.	TCP/IP stack policies are unchanged.	All policies are removed from the TCP/IP stack.
Policy Agent termination (NOPURGE defined)	TCP/IP stack policies are unchanged.	TCP/IP stack policies are unchanged.	TCP/IP stack policies are unchanged. Deleted policies are not removed from the TCP/IP stack.
Policy Agent update (FLUSH defined)	If there are any changed or deleted policies, then all policies are replaced in the TCP/IP stack.	Any changed policies are replaced in the TCP/IP stack, and then all deleted policies are removed from the TCP/IP stack.	Any changed policies are replaced in the TCP/IP stack, and then all deleted policies are removed from the TCP/IP stack.
Policy Agent update (NOFLUSH defined)	If there are any changed or deleted policies, then all policies are replaced in the TCP/IP stack.	Any changed policies are replaced in the TCP/IP stack, and then all deleted policies are removed from the TCP/IP stack.	Any changed policies are replaced in the TCP/IP stack. Deleted policies are not removed from the TCP/IP stack.

Table 43. How Policy Agent FLUSH and PURGE are used (continued)

Event	IPSec policies	Routing and ZERT policies	Other policies
Policy Agent refresh (FLUSH defined)	If there are any changed or deleted policies, then all policies are replaced in the TCP/IP stack.	If there are any changed or deleted policies, then all policies are deleted and reloaded into the TCP/IP stack.	If there are any changed or deleted policies, then all policies are deleted and reloaded into the TCP/IP stack.
Policy Agent refresh (NOFLUSH defined)	If there are any changed or deleted policies, then all policies are replaced in the TCP/IP stack.	If there are any changed or deleted policies, then all policies are deleted and reloaded into the TCP/IP stack.	Any changed policies are replaced in the TCP/IP stack. Deleted policies are not removed from the TCP/IP stack.

Rules:

- The TCP/IP stack results do not apply for policy client policies configured on the policy server.
- The PURGE and NOPURGE parameters have no effect on policy client policies configured on the policy server.

Result: When a TCP/IP stack is recycled, the result is the same as if the FLUSH parameter was specified; all active policies are reinstalled into the stack.

Switching between local and remote policies

If you dynamically switch from local policies to remote policies by adding the PolicyServer statement or a new PolicyType parameter within that statement, the FLUSH and PURGE parameters that are specified on the PolicyServer statement (or that are configured by default from the TcpImage statement) take effect, if the parameters are supported by the policy type.

Likewise, if you dynamically switch from remote policies to local policies by removing the PolicyServer statement or a PolicyType parameter from within that statement, the FLUSH and PURGE parameters that are specified on the *xxx*Config statement (or that are configured by default from the TcpImage statement) take effect, if the parameters are supported by the policy type.

When the NOFLUSH parameter is used due to one of these dynamic switches, the result is that both the local and remote policies exist in the configuration; existing policies are not deleted when NOFLUSH is in effect, as shown in Table 43 on page 828.

The following examples show how switching between local and remote policies works:

- Switching from local IDS to remote IDS policies:
  1. The TcpImage statement is configured with the FLUSH parameter.
  2. The IDSConfig statement is not configured with the FLUSH or NOFLUSH parameters, so the TcpImage FLUSH value is used.
  3. The local IDS policies are read and installed.
  4. The PolicyServer statement is added with a PolicyType parameter for IDS that specifies the NOFLUSH value.
  5. The remote IDS policies are retrieved and installed.
  6. Because the NOFLUSH parameter is in effect (from the PolicyServer statement), the local IDS policies are not deleted; both the local and remote IDS policies exist.
- Switching from remote AT-TLS to local AT-TLS policies:
  1. The TcpImage statement is configured with the NOFLUSH parameter.
  2. The TTLSConfig statement is configured with the FLUSH parameter.

3. The PolicyServer statement is configured with a PolicyType parameter for AT-TLS that specifies the FLUSH value.
4. The remote AT-TLS policies are retrieved and installed.
5. The PolicyType parameter for AT-TLS is removed from the PolicyServer statement.
6. The local IDS policies are read and installed.
7. Because the FLUSH parameter is in effect (from the TTLSConfig statement), the remote AT-TLS policies are deleted; only the local policies exist.

Result: Because the IPSec , Routing, and ZERT policy types always use the FLUSH value, the local and remote policies never exist at the same time.

Verifying that policies are correctly defined and functioning properly

To verify that policies are correctly defined and functioning properly, consider the following points:

- Are the policies defined correctly to the LDAP server?

  See the documentation appropriate for the LDAP server which you are using. LDAP servers usually allow you to install multiple files (LDIF), each containing different objects in the LDAP hierarchy. Structural objects higher in the directory tree must be installed before objects that are contained below them. Check for any error messages as each LDIF is installed. Some LDAP servers interpret two consecutive blank lines as end of file. Ensure that all of the objects in the LDIF have been installed by the LDAP server.

- Are the policies defined correctly to the Policy Agent?

  When starting the Policy Agent, first check for any error messages issued to the console. Message EZZ8434I indicates something is wrong with the Policy Agent environment. Message EZZ8438I indicates a syntax or semantic error in the policy definitions. Messages EZZ8439I and EZZ8440I indicate a problem with the LDAP server configuration or the server itself. For more information on diagnosing Policy Agent problems, see z/OS Communications Server: IP Diagnosis Guide. Use the UNIX **pasearch** command to display policy definitions. The output from this command indicates whether or not policy rules are active, and shows the parsed results of the policy definition attributes. One thing to note is that the Policy Agent is designed to ignore unknown attributes, so misspelled attributes will result in default values being used. The **pasearch** output can be used to verify that policies are correctly defined.

Chapter 15. Quality of service

Applications and users of IP networks have different requirements for the service they receive from those networks. A network that treats all traffic as best effort does not meet the needs of such users. *Service differentiation* is a mechanism to provide different service levels to different traffic types based on their requirements and importance in an enterprise network. For example, it might be critical to provide Enterprise Resource Planning (ERP) traffic better service during peak hours than that of FTP or web traffic. The overall service provided to applications or users, in terms of elements such as throughput and delay, is termed *Quality of service* (QoS). Network service providers that need to provide different QoS levels express their business goals in *Service Level Agreements* (SLAs). There are two types of service in IP networks that relate to QoS. The first is *Differentiated Services*, which provides QoS to broad classes of traffic or users, for example all outbound web traffic accessed by a particular subnet. The second is *Integrated Services*, which provides end-to-end QoS to an application, by reserving resources along a data path. For z/OS Communications Server, Integrated Services is largely provided by the RSVP Agent, which implements the Resource ReserVation Protocol.

Workload distribution also relates to QoS, in terms of the throughput and delay characteristics of a given server in a sysplex. The ability to dynamically monitor server performance and affect sysplex workload distribution is an important part of the overall QoS of a sysplex. Also important is the ability to limit the set of target systems considered for sysplex routing based on network selection criteria, such as source subnet.

Differentiated Services policies

Policies for Differentiated Services (DS) are used to select and control DS traffic for selected IP servers, such as FTP server traffic. The policy administrator selects the IP traffic to be controlled by defining policy rules. These policy rules include several attributes that can be specified to identify the traffic to be managed. These attributes fall into 2 categories, general attributes and application specified attributes. General attributes can be used to identify the IP traffic of most IP applications using a variety of information, such as:

- The source or destination IPv4 or IPv6 addresses or subnets
- The source and destination ports used by the application
- The IP protocol the application is using (TCP or UDP)
- The network interface selected for the outgoing traffic
- The jobname of the application

Application specified attributes allow policy administrators to identify outgoing application IP traffic based on information that is provided and defined by an application. For example, the IBM HTTP Server Powered by Domino® provides the TCP/IP stack with the URI (Universal Resource Identifier) associated with any outgoing data being sent to a client. This allows the policy administrator to define rules that identify traffic related to specific URIs and policy actions with unique DS controls for this traffic. For example, an installation can define a policy that specifies preferential treatment of outgoing traffic related to the servicing of any URIs beginning with `/product/placeorder`. For more information on defining policy rules for the IBM HTTP Server Powered by Domino based on URIs, see *z/OS HTTP Server Planning, Installing, and Using* and the policy configuration file topic in z/OS Communications Server: IP Configuration Reference.

Any IP application using the TCP protocol can provide application specified attributes using extensions to the sendmsg() socket API. For more information, see the programming interfaces appendix in the z/OS Communications Server: IP Programmer's Guide and Reference. Application provided attributes can be specified in 2 forms:

- Application defined data: This allows applications to provide free-form text data that can be used to classify the application's outgoing traffic in terms that should be familiar to the application's administrator (for example, URIs are provided by the IBM HTTP server).

© Copyright IBM Corp. 2000, 2022

- Application specified priority: This allows applications to associate an application priority on the outgoing IP traffic. This application priority in itself does not automatically cause the application's traffic to get preferential treatment. In order to make use of these application specified priorities the policy administrator needs to define policy rules that map these priorities to policy actions that will govern the outgoing traffic of each priority level.

Applications can pass both application defined data and application specified priorities to the TCP/IP stack. When both are specified, the administrator is free to use either or both criteria in their service policy rules. However, it is strongly recommended that any policy rules defined using the application specified attributes should also include at least one general attribute that uniquely identifies the application instance. For example, when defining rules for the HTTP server using URIs, you can help further identify the application by specifying the source port for the server or the HTTP Server's jobname. This will help ensure that unauthorized applications cannot exploit policy actions intended for the HTTP Server.

Several aspects of connection and throughput control can be specified with DS policies, including the following specifications:

- TCP connection limits
- Maximum and minimum TCP connection rates, TCP maximum delay
- Committed access bandwidth (mean rate and peak rate) control/enforcement, also known as token bucket traffic shaping
- IPv4 type of service (ToS) byte or IPv6 traffic class setting, and mapping to System z Queued Direct I/O (QDIO) device priorities and VLAN user priorities.

The above DS service attributes are enforced by the TCP/IP stack in which the DS policies are installed. For additional information on the enforcement of these attributes, see z/OS Communications Server: IP Configuration Reference.

Token bucket traffic shaping is defined using the following parameters:

- DiffServInProfileRate is the average or mean rate that you want to transmit over time. For example, 256 kilobits per second.
- DiffServInProfileTokenBucket is the *burst* size. This is how much data is allowed to be sent from the application to TCP/IP and still be allowed to be transmitted at the mean rate. It is suggested, if the application is not policing itself, that this burst size be at least one second's worth of data. Otherwise, if the application is sending large amounts of data at one time to TCP/IP, TCP/IP will slow that application down via congestion windows, and the mean rate may not be achieved.
- DiffServInProfilePeakRate is the highest rate that is allowed to be transmitted for a shorter interval of time. For example, though a customer might want on average only 256 Kb of data per second, they might allow a peak of 512 Kb of data for 1/4 second. The peak rate is used to control the spacing of outbound packets on the transmission line. By having a smaller peak rate, there will be longer spacing between packets, and thus less burstiness of traffic and increased efficiency. Higher peak rates result in shorter spacing and increased burstiness, which can result in lower link usage. However, some applications may require it, such as real time or video data.
- DiffServInProfileMaxPacketSize is the amount of data that will be policed at the peak cell rate. For example, if the peak rate is 512 Kb per second, and the maximum packet size is 120 Kb, TCP/IP will allow only about 10 packets of size 1492 bytes to be transmitted every .23 seconds. Again, if an application is sending large amounts of data at one time to TCP/IP, TCP/IP will enforce the peak rate, and anytime more than 10 packets are sent within .25 seconds, TCP/IP will begin to slow this TCP connection. The peak rate can be achieved over a longer period of time if the maximum packet size is entered in larger multiples of packets. However, this will cause greater burstiness as described above. For example, if the maximum packet size is entered as 240 Kb, TCP/IP will allow 20 packets in a .23 second range before enforcing slowdown.

  Note that the peak rate cannot be enforced without mean rate policing. However, you can enforce mean rate without peak rate. Also, setting of these parameters depends on the type of applications and the network that carries it.

Integrated Services policies

Integrated Services (RSVP) policies are used to set limits on certain parameters requested by RSVP applications. These applications interact with the RSVP Agent to establish resource reservations along a network path, using the RSVP API (RAPI). The reservation requests are in the form of an entity known as a Traffic Specification, or Tspec, which consists of the following values:

- Token bucket mean rate (r)
- Token bucket depth (b)
- Peak rate (p)
- Minimum policed unit (m)
- Maximum packet size (M)

RSVP policies can be used to limit the values requested for (r) and (b), as well as limiting the total number of RSVP reserved flows. The RSVP service attributes are enforced by the RSVP agent which gets RSVP policies from the Policy Agent. For additional information on the enforcement, see z/OS Communications Server: IP Configuration Reference or RFC 1363.

Sysplex distributor policies

Sysplex distributor (SD) policies are used to specify a set of SD target nodes for a given set of traffic. For example, all traffic destined to a given port or application from a specified subnet can be assigned one group of SD target nodes, while traffic for the same port or application from another subnet can be assigned to a different group of target stacks. These policies be can used in conjunction with other sysplex distributor controls to assist in load balancing. For more information, see “Policy interactions” on page 485.

QoS-specific Policy Agent functions

In addition to supporting the various types of policies, the Policy Agent performs functions related to the sysplex distributor. The Policy Agent can be configured to collect network QoS performance data relevant to SD on behalf of policies defined for a target port or application, and assign a default QoS weight fraction to such policy traffic. This weight is then used by SD (in conjunction with weights assigned by the Workload Manager) to assist in load balancing decisions. This function is performed by the Policy Agent on SD target nodes within the sysplex.

The Policy Agent also supports load distribution by service level. Performance data is kept for each Policy Action (service level) that a target's DVIPA port or application supports. A Policy Action weight fraction is generated. If available, this weight is used (instead of the default QoS weight fraction) in conjunction with the Workload Manager weight to assist in load distribution decisions for traffic assigned to this service level. If the Policy Action weight fraction is unavailable, the sysplex distributor will continue to use the default QoS weight fraction.

Another function related to policy performance is the performance collection function. When so configured, the Policy Agent collects policy performance data from the stack and caches it. This performance data is then made available to user applications through the Policy API (PAPI), for near real-time performance monitoring applications. The data are also optionally logged to a performance log file for offline performance monitoring. Sample C applications are provided to show how to use the PAPI interfaces to access performance data, and how to access and read the performance log file.

Policy performance data collected is affected by the FLUSH or NOFLUSH parameter on the Policy Agent TcpImage statement that defines the corresponding stack that is collecting the data. When FLUSH is specified, policies are deleted at the following times:

- When a new TcpImage statement is processed for the first time, including Policy Agent starting. This should not be a concern in most cases.
- When a MODIFY REFRESH command is entered.

As a result, all previously collected metrics start again from 0 when the policies are reinstalled. Conversely, policies are never deleted when NOFLUSH is specified, so performance metrics are never reset to 0.

Sysplex distributor policy performance monitoring and policy performance collection are similar in some respects but distinctly different in others:

- Sysplex distributor policy performance monitoring is actively performed by the Policy Agent. This performance monitoring is used only to assist with load balancing in a sysplex distributor environment.
- Policy performance collection is performed without regard to whether or not the Policy Agent is running in a sysplex distributor environment. Also, the Policy Agent does not actively participate in performance monitoring, only making the performance data available to user applications that perform the actual monitoring.

Policy performance data might not immediately change when changes are made to policy definitions. Some of the performance metrics are average values, and some are smoothed over several sampling intervals. As a result, when making changes to policies, some period of time will need to elapse before a new steady state is achieved.

Another function supported by the Policy Agent is to map IPv4 type of service (ToS) byte or IPv6 traffic class values to outbound interface priority values for outbound traffic. The ToS byte is also referred to as the Differentiated Services (DS) byte as an alternative definition (see RFC 2474). Note that outbound interface priority values are meaningful only for QDIO interfaces. A set of mappings can be defined to cover various ToS byte or traffic class values and map them to an appropriate interface priority. All outbound packets over the associated interfaces with a given ToS byte or traffic class value will then be assigned the corresponding priority value. ToS byte or traffic class values can also be mapped to Virtual LAN (VLAN) user priorities for propagation over LANs directly connected through the OSA-Express feature.

Note: Coding the virtual LAN (VLAN) user priority causes a frame to be sent out based on the IEEE 802.1Q specification, which establishes a standard method for tagging Ethernet frames with VLAN priority and membership information. Specifically, a VLAN priority-tagged frame is used to convey packet priority to the switches; it has a value of NULL for VLANID. A full VLAN-tagged frame contains both the priority and non-null VLANID. If you have switches in your network that do not support the IEEE 802.1Q standard or that are not properly configured for these types of frames, the frames might be dropped by the switch.

Sysplex distributor policy performance monitoring configuration

Before activating the sysplex distributor policy performance monitoring function, see "Policy interactions" on page 485 for information on workload balancing and policy interactions with sysplex distributor.

The following example illustrates how to activate the policy performance monitoring function for sysplex distributor.

Note: This function is activated on SD target servers and is used to monitor the performance of outbound traffic being serviced by the target servers. The goal is to detect TCP traffic that exceeds defined thresholds for dropped packets or time-outs, and derive a default QoS weight fraction for the target server. This default QoS weight fraction is then used to reduce the WLM weight assigned to the target servers, so that the SD distributing stack can take QoS performance into account.

The following statements apply to the example in this topic:

- The policy performance monitoring sampling interval is 60 seconds.
- Policy Agent assigns a loss ratio weight fraction of 25% when the TCP loss ratio (dropped packets to total packets) starts to exceed 2%.
- The loss ratio weight fraction is increased to 50% when the loss ratio starts to exceed 4%, continuing in this manner up to the maximum loss ratio weight fraction of 95%.
- In a similar manner, a TCP timeout weight fraction of 50% is assigned when the timeout ratio starts to exceed 5%, increasing up to a maximum timeout weight fraction of 100%.

- The loss ratio weight fraction and TCP timeout weight fraction are added together to form a single default QoS weight fraction for the target server, up to a maximum of 100%. When the Traffic Regulation policy connection limit reaches constrained threshold (90%), the default QoS weight fraction is set to 100% and forces SD to route requests to other target nodes with better routing weights.
- The default QoS weight fraction is used at the SD distributing stack to reduce the WLM weight. For example, if the WLM weight is 40, a weight fraction of 50% results in the weight being reduced to 20.
- The traffic to be monitored must be represented by at least one Differentiated Services policy defined for the target application (in this example a policy is defined for Telnet).
- An additional Policy Action weight fraction is calculated for a target's DVIPA/Port if there are any active connections to the target using that service level.

  The Policy Action weight fraction is calculated as the largest of three fractions:

  – The number of active connections to this target DVIPA/Port will be compared with the maximum connections allowed for this Policy action.

    - When the number of active connections reaches 50% of maximum connections, then the Policy Action weight fraction will be set to MAX (50%, current calculated value).
    - When the number of active connections reaches 65% of maximum connections, then the Policy Action weight fraction will be set to MAX (85%, current calculated value).
    - When the number of active connections reaches 80% of maximum connections, then the Policy Action weight fraction will be set to 100%.

  – The throughput rate for this timer interval will be calculated and compared to the DiffServ mean rate of this Policy action. If the throughput rate is greater than 85% of the DiffServ mean rate, the average throughput rate per connection will be calculated. If the throughput rate per connection is less than the DiffServ min rate, the minimum throughput requirement per connection is not being met and the Policy Action weight fraction will be set to 100%.

  – The default QoS weight fraction.
- Only one policy rule and policy action are defined here.

  As a result, only Telnet QoS performance information is monitored by the Policy Agent for sysplex distributor to route incoming Telnet connections to this target node relative to other target nodes which presumably can also accept Telnet requests.

```
PolicyPerfMonitorForSDR  enable
{
   samplinginterval 60
   LossRatioAndWeightFr 20 25
   TimeoutRatioAndWeightFr 50 50
   LossMaxWeightFr 95
   TimeoutMaxWeightFr 100
   MaxConnWeightFr 50 65 80
}
```

```
policyAction telnetGold
{
MinRate 500 # Provide minimum rate of 500 Kbps.
OutgoingTOS 10100000 # the TOS value of outgoing telnet packets.
}

policyRule targetelnet
{
ProtocolNumberRange 6
SourcePortRange 23
policyactionreference telnetGold
}
```

Policy performance collection configuration

The following example shows how to activate the policy performance collection function. Policy performance data for all active policies is maintained by the TCP/IP stack. This function allows this

data to be collected and made available for policy performance monitoring applications. The following statements apply to the example in this topic:

- Performance data is collected for both rules and actions. Action data is an aggregate of the data for the rules that refer to the action. For policies that have a single action per rule, the performance data will be the same for both the rule and action.
- The default minimum sampling interval is 30 seconds. This is the minimum value accepted for the acceptableCachedTime parameter on the PAPI papi_get_perf_data() function that gets performance data.
- Performance data will be logged to the file /u/user10/perflog, based on a sampling interval of 60 seconds.
- The number of performance log files maintained is 10, each of which is the default 300 kilobytes in size. In this example, the log files will be named assuming a stack name of TCPCS:

```
/u/user10/perflog.TCPCS
/u/user10/perflog.TCPCS.1
/u/user10/perflog.TCPCS.2
⋮
/u/user10/perflog.TCPCS.9
```

- Each performance data record is 232 bytes, so a file size of 300 kilobytes can contain 1324 records. Because 10 files are maintained, the total set of log files can contain 13240 records. Assuming that 50 policy rules and actions exist in the configuration, this means that the set of log files will wrap in approximately 4.4 hours, according to the following formula:

```
number of records (13240) / number of policies (50) = number of refresh cycles (264)

number of refresh cycles (264) * refresh interval in minutes (1) = 264 minutes worth of data
```

```
PolicyPerformanceCollection Enable
{
DataCollection           Rule Action
LogSamplingInterval      60
PerformanceLogFile       /u/user10/perflog
NumberOfLogFiles         10
}
```

IPv4 type of service or IPv6 traffic class mapping configuration

There are two mappings provided by the SetSubnetPrioTosMask statement:

- IPv4 type of service (ToS) or IPv6 traffic class to device priority

  Quality of service (QoS) support in z/OS Communications Server allows the IPv4 ToS byte, also known as the Differentiated Services (DS) field, or the IPv6 traffic class to be set for outbound IP packets according to defined policies managed by the z/OS Communications Server UNIX Policy Agent. When IP packets are sent out over QDIO devices, the ToS/DS or traffic class value is mapped to a QDIO priority value. Device priority values are 1-4, where 1 is the highest priority.
- IPv4 ToS or IPv6 traffic class to VLAN user priority

  ToS/DS or traffic class values can be mapped to user priorities for directly attached LANs using the OSA-Express feature in QDIO mode. VLAN user priority values are 0-7, where 7 is the highest priority. This allows assigned user priorities to be propagated through such networks, resulting in no loss of priority information for data being served by z/OS.

See z/OS Communications Server: IP Configuration Reference for more detail on these statements.

The following example shows a mapping of various ToS byte or traffic class values to associated interface priority values. Note that the mapping can be applied to individual interfaces or all interfaces:

- The first example defines a mapping for a specific interface. Note that the specified interface must be a valid interface specified in the HOME list. The second example shows a different mapping for all other interfaces.

- The subnet mask defines the bits in the ToS byte or traffic class that are significant. These examples use the leftmost 3 bits.
- The first example shows a set of mappings defining the complete set of ToS byte or traffic class values and the device and VLAN user priorities to be assigned for each value.
- The second example shows a set of mappings defining the complete set of ToS byte or traffic class values and the device priority to be assigned for each value.

```
SetSubnetPrioTosMask
{
  SubnetAddr         10.10.1.5
  SubnetTosMask      11100000
  PriorityTosMapping 1 11100000 7
  PriorityTosMapping 1 11000000 7
  PriorityTosMapping 2 10100000 6
  PriorityTosMapping 2 10000000 5
  PriorityTosMapping 2 01100000 5
  PriorityTosMapping 3 01000000 3
  PriorityTosMapping 4 00100000 2
  PriorityTosMapping 4 00000000 0
}
SetSubnetPrioTosMask
{
  SubnetTosMask      11100000
  PriorityTosMapping 1 11100000
  PriorityTosMapping 1 11000000
  PriorityTosMapping 1 10100000
  PriorityTosMapping 1 10000000
  PriorityTosMapping 2 01100000
  PriorityTosMapping 2 01000000
  PriorityTosMapping 3 00100000
  PriorityTosMapping 4 00000000
}
```

Options for configuring QoS

You configure QoS using a set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the QoS policy for each TCP/IP stack. In a complex environment, this file can become large. For this reason, there are two alternatives for creating the Policy Agent files.

Option 1: Use the IBM Configuration Assistant for z/OS Communications Server

The IBM Configuration Assistant for z/OS Communications Server, an optional GUI-based tool, provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the Policy Agent files.

The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

Through a series of wizards and online help panels, you can use the IBM Configuration Assistant for z/OS Communications Server to create QoS configuration files for any number of z/OS images with any number of TCP/IP stacks per image. Using the IBM Configuration Assistant for z/OS Communications Server, there are four types of reusable objects:

- Traffic descriptors that define the local application by describing the TCP traffic with ports or identifying the application using its job name.
- Priority levels that define the level of network priority or type of service (ToS).
- Traffic shaping levels that define settings to enforce specific traffic thresholds.
- Requirement maps that map traffic descriptors to priority levels and traffic shaping levels. A single requirement map should contain a complete set of QoS requirements that will govern the level of service for multiple IP traffic types.

For each TCP/IP stack, you select a requirement map that provides QoS for the stack.

The IBM Configuration Assistant for z/OS Communications Server comes with a number of IBM-supplied traffic descriptors, priority levels, traffic shaping levels and requirement maps that are easily applied, or you can use the IBM-supplied definitions as the basis for your own set of reusable objects.

The IBM Configuration Assistant for z/OS Communications Server can dramatically reduce the amount of time that is required to create QoS policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS, using the GUI can help you ensure that you have a consistent and easily manageable interface for implementing QoS.

This information primarily describes option 2, manual configuration. However, if you are using the IBM Configuration Assistant for z/OS Communications Server, reading this information will help you understand security concepts and the relationship between Policy Agent and QoS function.

Option 2: Manual configuration

You can manually create the QoS policy configuration files by coding all of the required statements in a file. There are a large number of configuration options provided by QoS policy statements that permit advanced users to carefully fine-tune QoS policy on a per-stack basis. This information describes the procedure for creating a QoS policy by manually creating and editing the configuration files. For details about the QoS policy statements, see z/OS Communications Server: IP Configuration Reference.

Specifying the QoS configuration file based on Policy Agent role

The Policy Agent can act as a policy server, a policy client, or neither. For more information on these different roles, see “Policy types and infrastructure overview” on page 787. Regardless of which option is used to configure QoS policies, the resulting configuration files need to be specified using different statements, depending on the role of the Policy Agent.

- If you are using the Policy Agent as a policy client that retrieves QoS policies from the policy server, specify the configuration files using the DynamicConfigPolicyLoad statement on the policy server.
- If you are using the Policy Agent as a policy client, but the policy client does not retrieve QoS policies from the policy server, specify the configuration files using the QOSConfig statement on the policy client, or configure the QoS policies directly in the image file specified on the TcpImage statement.
- If you are not using a policy client/policy server environment, specify the configuration files using the QOSConfig statement on the single Policy Agent, or configure the QoS policies directly in the image file specified on the TcpImage statement.

When specifying configuration files, keep in mind where the files should exist, based on the role of the Policy Agent.

Defining policies in a Policy Agent configuration file

Configure the following statements in the configuration file to define policies:

- PolicyAction
- PolicyRule

See z/OS Communications Server: IP Configuration Reference for more information about these statements.

The following subtopics contain examples of these tasks.

Note: These examples are for illustrative purposes only. The policies deliberately use a wide variety of attributes, and they do not necessarily represent real-world usage. Some examples show continued and indented statements that were modified to fit within the margin and therefore are not an actual representation of correct syntax.

Differentiated Services policy examples

The goal of this Differentiated Services policy is to map a subset of the traffic outbound from an FTP server.

This policy is identified as a Differentiated Services policy by the PolicyScope DataTraffic attribute on the PolicyAction statement, as well as the use of several DS-only attributes.

The following statements apply to the example in this topic:

- The policy rule selects traffic originated by ports in the range 20-21 (FTP outbound data connection uses port 20) from the source address 200.50.23.11.
- The policy rule is active on weekdays between 6 a.m. and 10 p.m. local time, between the dates 7/1/2000 and 7/1/2005..
- The policy action specifies that the ToS byte be set to '10000000' for traffic that conforms to this policy.
- The action establishes a *token bucket* traffic conditioner with a mean rate of 256 kilobits per second, a peak rate of 512 kilobits per second, and a burst size of 64 kilobytes. Any traffic that exceeds these specifications will be sent as *best effort*, with an accompanying ToS byte of '00000000'.

```
PolicyRule                                diffServ
    {
       ProtocolNumberRange                    6
       SourceAddressRange                     200.50.23.11
       SourcePortRange                        20-21
       PolicyActionReference                  tokenbucket
       PolicyRulePriority                     10
       ConditionTimeRange                     20000701000000:20050630235959
       DayOfMonthMask                         11111111111111111111111111111111
       DayOfWeekMask                          0111110
       TimeOfDayRange                         06:00-22:00
    }
    PolicyAction                              tokenbucket
    {
       PolicyScope                            DataTraffic
       OutgoingTOS                            10000000
       DiffServInProfileRate                  256       # 256 Kbps
       DiffServInProfileTokenBucket           512       # 512 Kbits
       DiffServInProfilePeakRate              512       # 512 Kbps
       DiffServInProfileMaxPacketSize         120       # 120 Kbits
       DiffServOutProfileTransmittedTOSByte   00000000
       DiffServExcessTrafficTreatment         BestEffort
    }
```

The goal of this policy is to ensure that outgoing data that match the specified attributes will be assigned a QoS service level defined in action "interactive1".

The following statements apply to the example in this topic:

- This rule will match traffic only on TCP connections (protocol 6) with a source port of 80 (i.e. HTTP server) and application defined data beginning with the string "/catalog".
- Because we are dealing with HTTP traffic, this rule is basically indicating that all outgoing traffic associated with a URI that begins with "/catalog" should be managed using the DS characteristics specified in the "interactive1" policy action.

```
PolicyRule          web-catalog       # web catalog traffic
{
     protocolNumberRange  6
     SourcePortRange    80
     ApplicationData /catalog
     policyActionReference  interactive1

}

PolicyAction      interactive1
{
    policyScope DataTraffic
    outgoingTOS 10000000
}
```

RSVP policy example

The goal of this RSVP policy is to establish limits on resource reservations requested by RSVP applications using the RSVP API (RAPI) interface. The policy is identified as an RSVP policy by the PolicyScope attribute on the PolicyAction statement, as well as the use of RSVP-only attributes.

The following statements apply to the example in this topic:

- The policy rule selects traffic from source ports in the range 8000 to 8001, with a protocol ID of 6 (TCP).
- The DataTraffic policy action specifies that the ToS byte be set to 01100000 for differentiated services traffic that conforms to this policy. Essentially, any traffic sent by the target application without an RSVP reservation in place will use this policy action. After an RSVP reservation is in place, the RSVP action gets used.
- The RSVP policy action specifies that the ToS byte be set to 01100000 while an RSVP reservation is in place. It also limits the type of RSVP service requested by RSVP applications to Controlled Load. Applications requesting Guaranteed service are downgraded to using Controlled Load service. In addition, the action limits the mean rate and token bucket size to 50000 bytes per second and 6000 bytes, respectively. These values are requested by RSVP applications in the traffic specification, or Tspec.
- The action also limits the number of active RSVP flows that map to this policy to 10.

```
PolicyRule                              intserv
     {
        SourcePortRange                      8000 8001
        ProtocolNumberRange                  6
        PolicyActionReference                intserv1
        PolicyActionReference                intserv2
     }
     PolicyAction                            intserv1
     {
        PolicyScope                          DataTraffic
        OutgoingTOS                          01100000
     }
     PolicyAction                            intserv2
     {
        PolicyScope                          RSVP
        OutgoingTOS                          01100000
        FlowServiceType                      ControlledLoad
        MaxRatePerFlow                       400    # 50000 bytes/second
        MaxTokenBucketPerFlow                48     # 6000  bytes
        MaxFlows                             10
     }
```

Sysplex distributor policy example

The goal of this sysplex distributor policy is to limit the number of SD target stacks for inbound Telnet traffic. The policies are identified as SD policies by the ForLoadDistribution TRUE attribute on the PolicyRule statement. The corresponding policy on the target is also shown.

The following statements apply to the example in this topic:

- Separate policies are defined on the sysplex distributor distributing and target stacks.
- The policy rules select incoming Telnet connection requests.
- The selected target stack will be based on WLM information and QoS information if activated at the target stacks.
- The rule (disttelnet) is coded on the distributing stack to select inbound traffic destined to the Telnet server.
- The rule (targtelnet) is coded on the target stack to select outbound data from the Telnet server.
- If none of the specified target stacks is available to service incoming requests (either the node is down or the Telnet server is not active), then sysplex distributor will distribute the requests to any available target stack.

Result: If the OutboundInterface 0.0.0.0 statement (for IPv4) and the OutboundInterface :: statement (for IPv6) are not present, and the defined target stacks are not available, sysplex distributor rejects the request.

```
policyAction        telnetGold
    {
        MinRate         500       # Provide minimum rate of 500 Kbps.
        OutgoingTOS     10100000  # the TOS value of outgoing telnet packets.
        outboundinterface  129.100.11.1
        outboundinterface  129.100.21.1
        outboundinterface  129.200.12.1
        outboundinterface  129::1B0D:13F0
        outboundinterface  0.0.0.0
        outboundinterface  ::
    }

    policyRule           disttelnet
    {
        ProtocolNumberRange     6
        DestinationPortRange    23
        PolicyRulePriority      20
        policyactionreference   telnetGold
        ForLoadDistribution     TRUE
    }

    policyRule           targtelnet
    {
        ProtocolNumberRange     6
        SourcePortRange         23
        PolicyRulePriority      20
        policyactionreference   telnetGold
        ForLoadDistribution     FALSE
    }
```

Notes:

1. The ApplicationName attribute is valid only for a target rule and should not be coded on a distributor rule because the application name determined for inbound traffic (which is always the case on a distributor) will always be the stack's TCP jobname.
2. If you are using Telnet with multiple stacks in conjunction with the sysplex distributor, see Chapter 11, “Accessing remote hosts using Telnet,” on page 569 for more information.

Defining policies using LDAP

For information about defining QoS policies using LDAP, see “Defining QoS policies using LDAP” on page 1389.

RSVP

Resource ReSerVation Protocol (RSVP) is a protocol that provides a mechanism to reserve resources in support of Integrated Services. The z/OS UNIX RSVP agent provides the following services on behalf of applications that want to use Integrated Services:

- An RSVP API (RAPI) that allows applications to explicitly request RSVP services. Using RAPI, applications indicate their intent to send or receive data, describe the characteristics of the data traffic and request that RSVP reserve resources along the data path to provide a given QoS to one or more traffic flows. For more information about RAPI, see z/OS Communications Server: IP Programmer's Guide and Reference.
- Mapping of IP ToS settings to RSVP traffic, using policies defined for RSVP.
- Support for VIPA addresses as well as real IP addresses.
- Communication with other RSVP agents on hosts and routers in the network to communicate application resource reservation requests.

Network administrators can use the z/OS UNIX Policy Agent to define RSVP-specific policies. These policies can be used to limit the parameters of application-requested resource reservations, provide

ToS mappings for RSVP traffic, and limit the number of traffic flows that can use RSVP services simultaneously.

RSVP is designed to be implemented on both end systems (hosts) and routers. Different functions are provided by RSVP in these two environments. The z/OS RSVP agent is supported as a host RSVP implementation only. It can communicate with router RSVP implementations, but is not itself supported as such. For more information about RSVP, see RFC 2205.

Configuring the RSVP agent

To configure the RSVP agent, update the configuration file to specify RSVP agent operational parameters using the LogLevel, TcpImage, Interface and RSVP statements. See z/OS Communications Server: IP Configuration Reference for detailed information about the statements.

To start the RSVP agent, you must first authorize the RSVP Agent using the security product. See SEZAINST(EZARACF) for SAF considerations for started tasks.

The following example is an RSVP configuration file. This example:

- Runs the RSVP Agent on the stack selected using the standard resolver search order, because a TcpImage statement is not configured.
- Disables RSVP processing on interface 10.11.12.13, while enabling it for all other interfaces.
- Disables traffic control on interface 200.1.1.1. This means that no reservations will be made on this interface.
- Allows a maximum of 50 active RSVP flows per interface.

```
Interface 10.11.12.13 Disabled
{}
Interface 200.1.1.1 Enabled
{
TrafficControl Disabled
}
Interface Others Enabled
{}
Rsvp All Enabled
{
MaxFlows 50
}
```

Starting and stopping RSVP

RSVP can be started from the z/OS shell or as a started task.

The RSVP agent uses the following search order to locate the configuration file (highest priority is listed first):

- z/OS UNIX file or MVS data set specified by the -c startup option. The syntax for a z/OS UNIX file is '/dir/file', and the syntax for an MVS data set is "//'MVS.DATASET.NAME'".
- z/OS UNIX file or MVS data set specified with the RSVPD_CONFIG_FILE enviroment variable.
- /etc/rsvpd.conf z/OS UNIX file.
- 'hlq.RSVPD.CONF' MVS data set.

  Note: If this file is not present, RSVP is enabled on all network interfaces with default parameters.

When starting from the shell, note that the RSVP executable file is in the /usr/lpp/tcpip/sbin directory. There is also a link from the /usr/sbin directory. Make sure your path statement (in the profile) contains either the /usr/sbin or /usr/lpp/tcpip/sbin directory.

Use the S RSVPD command on an MVS console or SDSF to start RSVP as a started task. A sample procedure is shipped in member EZARSVPP in SEZAINST.

The /tmp/rsvpd.pid.*tcpname* is a temporary RSVP daemon pid file that the RSVP daemon creates. This file contains the process ID of the current invocation of the RSVP daemon.

Restrictions:

- If /tmp/rsvpd.pid.*tcpname* is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the RSVP daemon.
- If /tmp/rsvpd.pid.*tcpname* is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /tmp/rsvpd.pid.*tcpname* is stored cannot have write access to the directory. Additionally, write access to /tmp/rsvpd.pid.*tcpname* must be limited to the owning UID or group, for example, --w--w----permissions.

RSVP can be stopped using the cancel command (C RSVPD) or using the kill command in the z/OS shell. The following kill command with the TERM signal will enable RSVP to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

where *pid* is the RSVP process ID.

The RSVP process ID can be obtained using the following z/OS UNIX command:

```
ps  -A
```

It can also be obtained from the /tmp/rsvpd.pid.*tcpname* file. See z/OS Communications Server: IP Configuration Reference for more information.

SNMP Network SLAPM2 (nslapm2) performance monitor

The SLAPM2 subagent provides information about defined service policies and performance data for applications that are mapped to those policies. Statistics are retrieved by this subagent and monitored for possible Network SLAPM2 performance deviations. For more information about the Network-SLAPM2 MIB, see usr/lpp/tcpip/samples/slapm2.mi2.

Configuring the Network SLAPM2 subagent

The z/OS CS Network SLAPM2 subagent allows network administrators to retrieve data and determine if the current set of Network SLAPM2 policy definitions are performing as needed or if adjustments need to be made. Before starting the Network SLAPM2 subagent, some basic operational characteristics of the Network SLAPM2 subagent need to be configured.

Procedure

Perform the following steps to configure the Network SLAPM2 subagent:

1. Configure the Network SLAPM2 subagent security authorization.

   For the subagent to retrieve performance monitor data from the Policy Agent, the Network SLAPM2 subagent must have superuser authority or security product authority in the SERVAUTH class.

   These profiles can be defined by TCP/IP stack and policy type as follows, where *sysname* is the system name defined in the sysplex, *TcpImage* is the TCP name for policy information that is being requested, and *ptype* is the policy type (*QOS*) that is being requested:

   ```
   EZB.PAGENT.sysname.TcpImage.ptype
   ```

   Wildcarding is allowed on segments of the profile name.

   If SERVAUTH class is absent (not RACLIST) or profiles are absent for the subagent's request (TcpImage and policy type), permission is denied and data is not returned.

   If SERVAUTH class is active, profiles are present for the subagent's request (TcpImage and policy type), and an MVS user is defined for all profiles, permission is granted and data is returned.

   If SERVAUTH class is active, profiles are present for the subagent's request (TcpImage and policy type), and an MVS user is not defined for all profiles, permission is refused and data is not returned.

For the sample commands needed to create the profile name and permit users to it, see the EZARACF sample in SEZAINST.

2. Configure the Policy Agent as follows:
   - Configure QoS policy rules and QoS policy actions in Policy Agent. The Network SLAPM2 subagent keeps statistics only for active QoS policies. For details, see "Steps for configuring the Policy Agent" on page 808.
   - Configure PolicyPerformanceCollection in Policy Agent. The PolicyPerformanceCollection statement for rules needs to be enabled to retrieve performance monitoring information from Policy Agent for the Network SLAPM2 Subagent. For details on the PolicyPerformanceCollection statement, see "Policy performance collection configuration" on page 835.
3. Configure and start the SNMP agent. For details, see "Step 1: Configure the SNMP agent" on page 1245.

Starting and stopping the Network SLAPM2 subagent

Before you start the Network SLAPM2 subagent, the following applications need to be be started and initialized:

- Policy Agent - For details, see "Starting and stopping the Policy Agent" on page 823.
- SNMP agent - For details, see "Start the SNMP agent" on page 1255.

The Network SLAPM2 subagent can be started from the z/OS shell or as a started task.

- When starting from the shell:

  The Network SLAPM2 subagent executable file (nslapm2) is in /usr/lpp/tcpip/bin. There is also a link from /bin. Make sure your PATH statement (in the profile) contains either /bin or /usr/lpp/tcpip/bin.

  The Network SLAPM2 subagent requires access to one or more DLLs at run time. The LIBPATH environment variable needs to be set to include the /usr/lib directory, which normally includes all the required DLLs.

  Export the LIBPATH environment variable before starting the subagent. This is best accomplished in /etc/profile or in .profile in the HOME directory. For example:

  ```
  export LIBPATH=/usr/lib
  ```

  The following command is an example:

  ```
  nslapm2 -d 3 -t 1800 -c special -P 5000
  ```

  The command above starts the Network SLAPM2 subagent with the following characteristics:

  – Connect to the SNMP agent using a community name of *special* and a port of *5000*.
  – The debugging level is set to *3*, to log the following debugging messages to syslogd:
    - Trace Network SLAPM2 subagent error and system console messages
    - Trace Network SLAPM2 subagent warning messages
  – The MIB table cache time is set to 30 minutes.
- When starting as a started task:

  Use the S NSLAPM2 command on an MVS console or SDSF. A sample procedure is shipped in member EZAPAGSB in SEZAINST.

  – Specify LIBPATH using the ENVAR parameter on the PARM statement in the started procedure. For example:

  ```
  // PARM=('POSIX(ON) ALL31(ON)',
  // 'ENVAR("LIBPATH=/usr/lib")/')
  ```

– Export the LIBPATH environment variable in a file specified with the STDENV DD statement. For example:

```
//STDENV DD PATH='/etc/nslapm2.env',PATHOPTS=(ORDONLY)
```

In the /etc/nslapm2.env file:

```
LIBPATH=/usr/lib
```

For more information on specifying run-time options, see z/OS Language Environment Programming Guide. For details on setting the LIBPATH environment variable, also see z/OS UNIX System Services Command Reference.

The /tmp/nslapm2.*tcpname*.pid is a temporary NSLAPM2 subagent pid file that the Network SLAPM2 subagent creates. This file contains the process ID of the current invocation of the Network SLAPM2 subagent.

Restrictions:

- If /tmp/nslapm2.*tcpname*.pid is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the Network SLAPM2 subagent.
- If /tmp/nslapm2.*tcpname*.pid is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /tmp/nslapm2.*tcpname*.pid is stored cannot have write access to the directory. Additionally, write access to /tmp/nslapm2.*tcpname*.pid must be limited to the owning UID or group, for example, --w--w----permissions.

The Network SLAPM2 subagent can be stopped using the stop command (P NSLAPM2), or using the kill command in the z/OS shell. For example, the following kill command with the TERM signal, where *pid* is the nslapm2 process ID, enables the Network SLAPM2 subagent to clean up resources properly before terminating itself:

```
kill -s TERM pid
```

The nslapm2 process ID can be obtained using the following z/OS UNIX command:

```
ps  -A
```

Verification

To verify that policies are correctly defined and functioning properly, consider the following points:

- Are the policies installed in the TCP/IP stacks?
- Is the expected traffic mapping to the correct policies?
- Are the sysplex distributor policy functions working correctly?
- Does anything need to be tuned?

The following subtopics provide more details about these considerations.

Verifying that the policies are installed in the TCP/IP stacks

Use the Netstat SLAP/-j command to display QoS policy statistics. This command displays statistics for only active (installed) QoS policies, so it can be used to verify the correct policies are installed, even if all the statistics are 0. Because the Policy Agent can install policies on multiple stacks, issue this command on each stack to verify the correct set of QoS policies is installed.

Verifying that the expected traffic is mapping to the correct QoS policies

While connections are active, use the Netstat ALL/-A command to display details about the active connections. One piece of information displayed is the policy rule name. If this name is blank, then the traffic is not mapped to any active rule. Also, use the Netstat SLAP/-j command to display QoS

policy statistics. The output shows the time that each policy was last mapped to traffic, and accumulated statistics for each policy. Monitor these values over time to verify that new traffic is mapping as expected.

Note: The values displayed by the Netstat SLAP/-j commands can wrap around to 0. If some of the values do not seem correct (for example, total out bytes less than total out bytes in profile), then wrapping has probably occurred.

Verifying that the sysplex distributor policy functions are working correctly

To verify that the distributor is using the expected service levels when deciding how to distribute traffic to each DVIPA/Port target, use the Netstat VDPT/-O DETAIL command on the distributing stack. The following QoS related information will be displayed for each DVIPA/Port target:

- WLM weight unmodified by QoS
- Modified WLM/QoS aggregate weight, identified by *DEFAULT*
- Modified WLM/QoS service level weights, identified by service level name

To verify that active connections distributed to DVIPA/Port targets are using the expected service level, use the Netstat VCRT DETAIL command on the distributing stack. This will display the following policy related information:

- PolicyRule: the policy rule that the distributor used in selecting the policy action for this connection.
- PolicyAction: the policy action that this connection is currently using. If PolicyAction is specified by *NONE*, then the distributor is using the *DEFAULT* fraction to distribute this connection.

See z/OS Communications Server: IP Diagnosis Guide for more information.

Monitoring performance and tuning policies

Poor performance, such as low throughput, long response times, and so on, might be suddenly and consistently experienced by a certain set of users or applications. Also, traps might be generated by the Network SLAPM2 subagent. When this happens, the problem might be the way the QoS policy is defined for the corresponding set of users or applications.

For example, the IPv4 ToS/DS or IPv6 traffic class value might be set incorrectly to a lower QoS level than is intended, for example, medium or low priority instead of high priority. It is important to remember that given a fixed amount of network resources, changing some traffic demand from a lower to higher QoS level will mean that other traffic demands will be affected. Therefore, use care to ensure that in attempting to meet one set of QoS requirements, different or worse problems do not result.

Another cause for poor performance might be in the way the bandwidth allocation defined via the DiffServ token bucket parameters, or TCP maxrate or minrate, is not adequate to accommodate the traffic demand. Yet another possibility might be that either network or the server capacity is not adequate to handle the traffic demand. This is evident when a majority of users or applications do not have their QoS requirements met. When this happens, the network planning process must be revisited.

For more information, see “Using the Network SLAPM2 MIB to monitor policies” on page 846.

Using pasearch

Use the **pasearch** command to display policy details. This command displays both active (installed in the stack) and inactive policies. Various parameters can be specified to filter the results, for example to display only policies for certain stacks, only QoS policies, only policy names, or only a single policy specified by name. See z/OS Communications Server: IP System Administrator's Commands for the complete syntax and sample output for **pasearch**.

Using the Network SLAPM2 MIB to monitor policies

The Network SLAPM2 subagent provides information about service policies and performance data for applications mapped to those policies through the slapm2PolicyRuleStatsTable.

Note: The Network SLAPM2 subagent can be used to monitor Differentiated Services policies.

slapm2PolicyRuleStatsTable
Provides statistics on a per policy rule basis.

The Network SLAPM2 subagent also supports performance monitoring with the slapm2PRMonTable object. Entries are created in the monitor table to establish the wanted criteria for monitoring. The following level of monitoring is provided:

Aggregate
Monitoring is performed based on the aggregate of all TCP or UDP applications that are mapped to one or more service policies.

Three types of monitoring are provided for measuring application performance:

TCP round-trip time
The current TCP round-trip time of applications are compared to the threshold values established in the monitor table entry. If the current rates exceed the high threshold or go below the low threshold, an SNMP trap is sent if traps are enabled.

TCP packets retransmit ratio
The current TCP packets retransmit ratios of applications are compared to the threshold values established in the monitor table entry. If the current rates exceed the high threshold or go below the low threshold, an SNMP trap is sent if traps are enabled.

Average accept queue delay
The current average accept queue delay of applications are compared to threshold values established in the monitor table entry. If the current rates exceed the high threshold or go below the low threshold, an SNMP trap is sent if traps are enabled.

For more details about how to make the various monitoring calculations, see the NETWORK-SLAPM2-MIB in the sample file slapm2.mi2 in the /usr/lpp/tcpip/samples directory.

When SNMP traps are enabled, and a *not achieved* trap is sent as described above, a corresponding *okay* trap is sent when the traffic again conforms to the boundaries established in the monitor table entry.

For example, suppose the slapm2PRMonTcpRttDelayHigh value is set to 2 seconds and the slapm2PRMonTcpRttDelayLow value is set to 1 second. If the TCP round-trip delay rises above 2 seconds, a *not achieved* trap is sent. If the TCP round-trip delay then drops below 1 second, an *okay* trap is sent to indicate the problem has been resolved. However, if the row becomes inactive before conforming to the established boundaries, an *okay* trap is never sent, because this removes monitoring for this entry.

The following traps are used to monitor table administration:

Policy deleted
A slapm2PolicyRuleDeleted trap is sent when an entry is deleted from the slapm2PolicyRuleStatsTable, if slapm2PolicyDeletedTrapEnabled is enabled(1).

Monitor deleted
A slapm2PolicyRuleMonDeleted trap is sent when a slapm2PRMonEntry is deleted, if the value of slapm2PolicyDeletedTrapEnabled is enabled(1).

Creating monitor table entries and enabling SNMP traps

Several MIB objects are used when establishing monitor table entries and when configuring whether and how often traps are sent. To establish monitor table entries, set the following MIB object variables. Most of these objects have default values, so you might be able to achieve the wanted monitoring using only a subset of the objects.

slapm2PRMonTcpRttDelayHigh, slapm2PRMonTcpRttDelayLow
Establishes the threshold values for the average TCP round-trip time. The minimum and maximum rates are in units of milliseconds.

slapm2PRMonTcpReXmitHigh, slapm2PRMonReXmitLow
Establishes the threshold values for the TCP packets retransmit ratio. The minimum and maximum rates are in tenths of a percent units.

slapm2PRMonTcpAcceptQDelayHigh, slapm2PRMonTcpAcceptQDelayLow
Establishes the threshold values for the average accept queue delay. The minimum and maximum rates are in units of milliseconds.

slapm2PRMonRowStatus
This object allows entries to be created and deleted in the slapm2PRMonTable.

In addition, the following MIB objects are used to control the generation of traps:

slapm2PRMonTrapEnable
Indicates whether slapm2PolicyRuleMonNotOkay and slapm2PolicyRuleMonOkay notifications should be generated for this conceptual row.

slapm2PRMonTrapFilter
The purpose of this object is to suppress excessive slapm2PolicyRuleMonNotOkay notifications. A monitored quantity must exceed its high threshold for the number of consecutive intervals indicated by this object for a notification to be generated. The length of the intervals is specified by the slapm2PolicyMonInterval object.

Creating the monitor table index

When you create monitor table entries, specify the appropriate index value. The index is composed of:

- slapm2PRMonOwnerIndex
- slapm2PolicyRuleIndex

The OwnerIndex is expressed in the following format, where *character* is in ASCII decimal form:

```
length.character.character...
```

For example, the value *u1* is expressed as 2.117.31. The PolicyRuleIndex that maps to the policy name value is the index into the slapm2PolicyRuleTable.

The Network SLAPM2 subagent creates an entry in the slapm2PolicyRuleTable to represent a policy rule. The index value for this entry is arbitrary and assigned by the subagent. Corresponding entries in the other MIB tables, including the monitor table, contain the index value that maps to the entry in the name table.

To assist you in creating the index for the monitor table entries, note that the index value used in the slapm2PolicyRuleStatsTable entries consist of the last value used in the monitor table index, namely the PolicyRuleIndex. Thus, you can walk through the policy statistics table using the following command:

```
osnmp -v walk slapm2PolicyRuleStatsTable
```

Then, cut and paste the index value from the PolicyRuleStatsTable and add an OwnerIndex of your choosing at the beginning of the index.

For the above example, the complete index using an OwnerIndex of *u1* is:

```
2.117.31.3
 |       +--- name table index value (PolicyRuleIndex)
 +----------- length + "u1" (OwnerIndex)
```

Monitor table examples

If you are going to change any of the monitor table object values for an existing table entry or row, you must take the row out of service to make the changes. To do this, set the value of slapm2PRMonRowStatus to 2. After your changes are made, set the row status to a value of 1 to put it back in service.

The following examples show how to create monitor table entries to monitor.

- This example assumes SNMP version 1 security and no SNMPD.CONF file.

  1. Enable traps. The snmptrap.dest file should contain the IP address and protocol of an entity to receive traps:

```
/etc/snmptrap.dest   contains: 9.67.191.5        UDP
/etc/pw.src          contains: public 0.0.0.0    0.0.0.0
```

In this example, use the osnmp command running in the background to receive traps:

```
osnmp trap > /tmp/trap.output &
```

2. Change status to notInService:

```
osnmp set slapm2PRMonRowStatus.index 2
```

3. Enable monitoring for slapm2PolicyRuleMonNotOkay and slapm2PolicyRuleMonOkay (traps):

```
osnmp set slapm2PRMonTrapEnable.index 1
```

4. If wanted, change default thresholds:
   - TCP round trip, where *l* is the lower boundary and *h* is the upper boundary:

     ```
     osnmp set slapm2PRMonTcpRttDelayLow.index l
     osnmp set slapm2PRMonTcpRttDelayHigh.index h
     ```

   - TCP retransmit ratio, where *l* is the lower boundary and *h* is the upper boundary:

     ```
     osnmp set slapm2PRMonTcpReXmitDelayLow.index l
     osnmp set slapm2PRMonTcpReXmitDelayHigh.index h
     ```

   - Accept Queue delay ratio, where *l* is the lower boundary and *h* is the upper boundary:

     ```
     osnmp set slapm2PRMonAcceptQDelayLow.index l
     osnmp set slapm2PRMonAcceptQDelayHigh.index h
     ```

5. Make row active:

```
osnmp set slapm2PRMonRowStatus.index 1
```

- Evaluate the following fields to determine why the slapm2PolicyRuleMonNotOkay trap was generated:
  - If the maxTcpRttDelayExceeded bit in the previous slapm2PRMonStatus is off, indicating below the high threshold, and the bit in the current slapm2PRMonStatus is on, indicating above the high threshold, this indicates that at the end of this monitor interval this is a rising quantity and the threshold has exceeded its high threshold. Evaluate the slapm2PRMonTcpRTTCurrentDelay to determine the average round-trip time over the most recent interval for all outgoing TCP packets affected by this policy rule.
  - If the maxTcpReXmitRatioExceeded bit in the previous slapm2PRMonStatus is off, indicating below the high threshold, and the bit in the current slapm2PRMonStatus is on, indicating above the high threshold, this indicates that at the end of this monitor interval this is a rising quantity and the threshold has exceeded its high threshold. Evaluate the slapm2PRMonTcpCurrentTcpReXmit to determine the TCP retransmit ratio over the most recent interval for all outgoing TCP packets affected by this policy rule.
  - If the maxAcceptQueueDelayExceeded bit in the previous slapm2PRMonStatus is off, indicating below the high threshold, and the bit in the current slapm2PRMonStatus is on, indicating above the high threshold, this indicates that at the end of this monitor interval this is a rising quantity and the threshold has exceeded its high threshold. Evaluate the slapm2PRMonAcceptQCurrentDelay to determine the smoothed average accept queue delay over the most recent interval for all flows affected by this policy rule.
- Evaluate the following fields to determine why the slapm2PolicyRuleMonOkay trap was generated:
  - If the maxTcpRttDelayExceeded bit in the previous slapm2PRMonStatus is on, indicating above the low threshold, and the bit in the current slapm2PRMonStatus is off, indicating below the low threshold, this indicates that at the end of this monitor interval this is a falling quantity and the threshold has fallen below its low threshold. Evaluate the slapm2PRMonTcpRTTCurrentDelay to

determine the average round-trip time over the most recent interval for all outgoing TCP packets affected by this policy rule.

- If the maxTcpReXmitRatioExceeded bit in the previous slapm2PRMonStatus is on, indicating above the low threshold, and the bit in the current slapm2PRMonStatus is off, indicating below the low threshold, this indicates that at the end of this monitor interval this is a falling quantity and the threshold has fallen below its low threshold. Evaluate the slapm2PRMonTcpCurrentTcpReXmit to determine the TCP retransmit ratio over the most recent interval for all outgoing TCP packets affected by this policy rule.
- If the maxAcceptQueueDelayExceeded bit in the previous slapm2PRMonStatus is on, indicating above the low threshold, and the bit in the current slapm2PRMonStatus is off, indicating below the low threshold, this indicates that at the end of this monitor interval this is a falling quantity and the threshold has fallen below its low threshold. Evaluate the slapm2PRMonAcceptQCurrentDelay to determine the smoothed average accept queue delay over the most recent interval for all flows affected by this policy rule.

Chapter 16. Intrusion detection services

It is becoming increasingly important to not just protect systems from attacks but to detect patterns of usage that might indicate impending attacks. Many attacks follow a sequence of information gathering, unauthorized access to resources (information, applications, storage), and denial of service. It can be difficult, or at times, impossible to determine the originator of denial of service attacks. Correlating information gathering activities with access violation might help identify an intruder before they succeed.

Intrusion detection services (IDS) provides the following support:

- Scan detection and reporting
- Attack detection, reporting, and prevention
- Traffic regulation for TCP connections and UDP receive queues

You can use IDS policies to specify event conditions and the actions to take for particular events. All IDS policies support logging events to a specified message priority level in syslogd, on the system console, or both. Most IDS policies support the following functions:

- Discarding packets when a specified limit is reached
- Writing statistical records to the INFO message level of syslogd at a specified time interval, and the option to write these records only when exceptional events have occurred
- Tracing all or part of the triggering packet to an IDS-specific CTRACE facility, SYSTCPIS

IDS assigns a correlator value to each event. All messages written to the system console and syslogd use this correlator, and records written to the IDS trace use this correlator. A single detected event can involve multiple packets; the correlator value identifies which messages and packets are related to each other. Each IDS policy has additional attributes that you can specify, either in conditions or in the action.

Scan policies

Scans are recognized when a single source IP address makes multiple attempts to gather information within a defined period of time. Although a scan is not harmful, many serious attacks, especially access violation attacks, are preceded by scans to gather information. Scans must use reliable source IP addresses and can be interesting events to monitor.

IDS support defines a scanner as a source host that accesses multiple unique resources (ports or interfaces) over a specified period of time. You can specify the number of unique resources (Threshold) and the time period (Interval) by policy. Two categories of scans are supported:

- Fast scan

  Many resources are rapidly accessed in a short time period (usually less than 5 minutes and program driven)
- Slow scan

  Different resources are intermittently accessed over a longer period of time (many hours). This could be a scanner trying to avoid detection.

The following scenarios are some examples of a scanner:

- Source host A has a program that loops through all low ports and tries to connect to each port on target host X (fast scan). Each port is considered a unique resource.
- Source host B manually sends a **ping** command to each interface on target host X, and then tries to access well-known ports on target host X (most likely a slow scan). Each interface accessed by a **ping** command is considered a unique resource, and each port accessed is considered a unique resource.

The following scenario is not considered to be an example of a scanner:

© Copyright IBM Corp. 2000, 2022

- Source host C starts 20 connections to port 23. Because these connections are to the same port, only one unique resource is accessed, and host C is not considered a scanner.

Scans like the one in the following scenario might not be detected by IDS:

- Source host E issues a **ping** command to IP addresses 9.1.1.1 through 9.255.255.255. Because host X collects data only for the ping commands that are directed to interfaces of host X, this event is not detected by host X as a scan. Network IDS might detect this as wide scan.

Scan policy provides the ability to specify the following conditions and actions:

- Fast scan time interval
- Slow scan time interval
- Fast scan threshold
- Slow scan threshold
- Exclude well-known legitimate scanners by using an exclusion list
- Specify a sensitivity level by port or port range (to reduce performance impacts)
- Provide notification of a detected scan by issuing a console message or a syslogd message
- Trace potential scan packets

The individual packets used in a scan are categorized as normal, possibly suspicious, or very suspicious. To control the performance impact and analysis load of scan monitoring, set the sensitivity level for potential scan events to high, medium, or low.

Table 44 on page 852 shows how the policy-specified sensitivity affects the counting of scan events. The event suspicion level is determined by the stack.

Table 44. Effect of sensitivity level on the counting of scan events at each suspicion level

Sensitivity (from policy)	Normal event	Possibly suspicious event	Very suspicious event
Low			counted
Medium		counted	counted
High	counted	counted	counted

To help reduce or eliminate false positive results, IDS allows policy-specified source IP addresses, subnet masks, and optionally, source port numbers, to be excluded from scan detection. For UDP and TCP port scans, you can limit scan detection to specific destination port ranges. You can specify the sensitivity level (high, medium, or low) for these port ranges.

Another way that IDS reduces false positive results is by counting only unique events from a specific source IP address within a scan interval. An event is considered unique if the IP protocol, destination IP address, and destination port (UDP or TCP) or type (ICMP) have not been encountered before during this scan interval.

IDS scan policy supports a fast scan interval and threshold, and a slow scan interval and threshold:

- Fast scan

  A fast scan is recognized if the number of unique events within the fast scan interval reaches the fast scan threshold.

- Slow scan

  A slow scan is recognized if the number of unique events within the slow scan interval reaches the slow scan threshold.

IDS counts scan events using an internal interval that is no greater than half of the fast scan interval. During an internal interval, after the number of unique events reaches the slow scan threshold, additional events are not counted. These additional events are traced if requested by policy using the trace data parameter in the action.

Restriction: When system resources are constrained, IDS might temporarily suspend scan detection.

Scan events are categorized by the following scan types:

- ICMP scans
- ICMPv6 scans
- UDP port scans
- TCP port scans

Any countable scan event is counted against the origin source IP address, and the total number of countable events from all categories is compared to the policy thresholds. When an origin source IP address has reached the fast or slow threshold defined by policy, the following actions are taken if you requested them using the notification options in the action:

- A notification is sent to the traffic regulation management daemon (TRMD) for logging to syslogd
- A console message is issued
- The packet is logged to the IDS packet trace

When IDS detects a scan for a particular source IP address, no further scan events are reported for that source IP address during the specified interval. The intervals and thresholds for fast and slow scans are global; one set of values applies to all event categories.

ICMP scans

Internet Control Message Protocol (ICMP) requests (Echo, Information, Timestamp, and Subnet Mask) are used to map network topology. Receipt of an ICMP request is classified as a normal, possibly suspicious, or highly suspicious event. Any request sent to a multicast or broadcast address is treated as a very suspicious event. Echo requests (ping) and Timestamp requests are common and are treated as normal events when they do not include the IP Options for Record Packet Route or Record Timestamp. Events are classified by the first matching entry in Table 45 on page 853:

Table 45. Classification of ICMP events

Event	Destination address	Event classification
Receive any ICMP request (Echo, Information, Timestamp, or Subnet Mask)	Multicast or broadcast	Very suspicious
Receive any ICMP request that is denied by Quality of Service (QoS) policy	Unicast	Normal
Receive Information Request or Subnet Mask	Unicast	Possibly suspicious
Receive Echo Request with IP Option Record Route or Record Timestamp	Unicast	Possibly suspicious
Receive Echo Request or receive Timestamp Request	Unicast	Normal

ICMPv6 scans

ICMPv6 provides ICMP function for IPv6. Table 46 on page 853 shows the ICMPv6 events that are considered for scan detection and the suspicion level associated with each event. Events are classified by the first matching entry in the table.

Table 46. Classification of ICMPv6 events

Event	Destination address	Event classification
Receive Echo Request	Multicast	Very suspicious

Table 46. Classification of ICMPv6 events (continued)		
Event	**Destination address**	**Event classification**
Receive Echo Request that is denied by Quality of Service (QoS) policy	Unicast	Normal
Receive Echo Request with Routing Header	Unicast	Possibly suspicious
Receive Echo Request without Routing Header	Unicast	Normal

UDP port scans

Because UDP is stateless, the stack cannot differentiate between a client port and a server port. A scanner that is sending messages to many ephemeral ports looks similar to a DNS server that is sending replies to many clients on ephemeral ports. You can specify the RESERVED keyword on the PORT or PORTRANGE statement in the TCP/IP profile to prohibit the use of a UDP port. Any datagram that is received for a prohibited port is treated as a very suspicious event. Any datagram that is received for a port that is not prohibited but is unbound is treated as a possibly suspicious event, and any datagram received for a bound port is treated as a normal event. You can also limit event generation to specific port ranges and destination addresses. UDP port scans apply to IPv4 and IPv6 packets. Events are classified by the first matching entry in Table 47 on page 854:

Table 47. UDP port event classification		
Socket state	**Event**	**Event classification**
Any state	Receive any packet that is denied by IP security filtering	Possibly suspicious
Use prohibited by RESERVED keyword	Receive any packet	Very suspicious
Unbound, use not prohibited	Receive any packet	Possibly suspicious; application could be temporarily down
Bound	Receive any packet	Normal

TCP port scans

Because TCP is a stateful protocol, many different events might be classified as normal, possibly suspicious, or very suspicious. The identified conditions are listed in Table 48 on page 854. You can use the RESERVED keyword on the PORT or PORTRANGE statement in the TCP/IP profile to prohibit the use of a TCP port. You can also limit event generation to specific port ranges and destination IP addresses. TCP port scans apply to IPv4 and IPv6. Events are classified by the first matching entry in Table 48 on page 854:

Table 48. TCP port event classification		
Socket state	**Event**	**Event classification**
Any state	Receive unexpected flags (for example, SYN+FIN)	Very suspicious
Any state	Receive standalone SYN that is denied by IP security filtering	Possibly suspicious
Use prohibited by RESERVED keyword	Receive standalone SYN	Very suspicious

Table 48. TCP port event classification (continued)		
Socket state	**Event**	**Event classification**
Unbound, use not prohibited	Receive standalone SYN	Possibly suspicious; application could be temporarily down
Listen	Receive standalone SYN that is denied by Quality of Service (QoS) policy	Normal
Listen	Receive standalone SYN	No event (classification deferred)
Half open connection	Receive ACK	Normal; connection handshake completed
Half open connection	Receive duplicate SYN	Normal; perhaps duplicate packet
Half open connection	Receive RST	Possibly suspicious; peer covering tracks
Half open connection	Final time out	Very suspicious; peer abandoned handshake
Any connected state	Seq# out of window	Normal; perhaps duplicate packet
Any connected state	Receive standalone SYN	Normal; perhaps peer reboot
Any connected state	Final timeout	Possibly suspicious; peer abandoned connection

Attack policies

An attack can be a single packet designed to cause a system to fail or hang. An attack can also consist of multiple packets designed to consume a limited resource, which causes a network, system, or application to be unavailable to its intended users (denial of service). You can use intrusion detection services (IDS) attack policy to activate attack detection for one or more categories of attacks independently of each other. In general, the types of actions that you can specify for an attack policy are event logging, statistics gathering, packet tracing, and discarding of the attack packets. Most attack checking is performed for inbound packets to a stack.

IDS includes the following categories of attacks:

- Malformed packet events

  Many attacks are designed to cause the protocol stack of a system to fail by providing incorrect or partial header information. These packets are always discarded when they are received, regardless of IDS policy. The source IP address is rarely reliable for these attacks. This type of attack applies to both IPv4 and IPv6 packets.

  You can use IDS policy to provide notification of malformed packet attacks.
- Inbound fragment events

  Many attacks are the result of fragment overlays in the IP or transport header. You can use this support to detect fragmentation overlays that change the data in a packet, including changes to the length of the packet. This type of attack applies to both IPv4 and IPv6 packets.

  You can use IDS policy to provide notification of suspicious fragments, and optionally to discard them.
- IP protocol restrictions

Although there are 256 possible values that can be specified in the protocol field of the IP header, only a handful are commonly used. You can protect your system against future attacks by prohibiting those values that you are not actively and intentionally using.

Restriction: This attack type applies to IPv4 packets only. The IPv6 next header restrictions attack type provides analogous function for IPv6 packets.

You can use IDS policy to provide notification of a packet with a restricted IP protocol value, as well as to discard the packet.

- IPv6 next header restrictions

  The IPv6 header and any subsequent extension headers include a next header field. The value in the next header field identifies the next header in the packet, either an upper layer protocol header (such as a TCP or UDP header) or an extension header (such as a fragmentation or routing header). Although there are 256 possible values that can be specified as a next header value, only a small number are commonly used. You can protect your system by prohibiting next header values in inbound packets that you are not actively and intentionally using.

  Restriction: This attack type applies to IPv6 packets only. The IP protocol restrictions and IP option restrictions attack types provide analogous function for IPv4 packets.

  You can use IDS policy to provide notification of an inbound packet with a restricted next header value, as well as to discard the packet.

- IP option restrictions

  As with IP protocols, there are 256 possible values that can be specified as IP options in received packets, with only a small number currently in common use. You can prevent misuse of values that you are not intentionally using. A check for restricted IP options is performed on all inbound packets, including those that are forwarded to another system.

  Restriction: This attack type applies to IPv4 packets only. The IPv6 destination option restrictions, IPv6 hop-by-hop option restrictions, and IPv6 next header restrictions attack types provide analogous function for IPv6 packets.

  You can use IDS policy to provide notification of a packet with a restricted IP option, as well as to discard the packet.

- IPv6 destination option restrictions

  Although there are 256 possible values that can be specified as IPv6 destination options in received packets, only a small number of values are currently defined. You can prevent misuse of values that you are not intentionally using.

  Restriction: This attack type applies to IPv6 packets only. The IP option restrictions attack type provides analogous function for IPv4 packets.

  You can use IDS policy to provide notification of a packet with a restricted value for an IPv6 destination option, as well as to discard the packet.

- IPv6 hop-by-hop option restrictions

  Although there are 256 possible values that can be specified as IPv6 hop-by-hop options in received packets, only a small number of values are currently defined. You can prevent misuse of values that you are not intentionally using. A check for restricted IPv6 hop-by-hop options is performed on all inbound packets, including those that are forwarded to another system.

  Restriction: This attack type applies to IPv6 packets only. The IP option restrictions attack type provides analogous function for IPv4 packets.

  You can use IDS policy to provide notification of a packet with a restricted value for an IPv6 hop-by-hop option, as well as to discard the packet.

- UDP perpetual echo

  Some UDP applications unconditionally respond to every datagram that is received. In some cases, such as Echo, CharGen or TimeOfDay, this is a useful network management or network diagnostic tool. In

other cases, an application might send error messages in response to incorrectly formed requests. If a datagram is inserted into the network with one of these applications as the destination and another one of these applications spoofed as the source, the two applications will respond to each other continually. Each inserted datagram results in another perpetual echo conversation between them. You can use IDS policy to define the application ports that exhibit this behavior. This type of attack applies to both IPv4 and IPv6 packets.

You can use IDS policy to provide notification of a perpetual echo packet, as well as to discard the packet.

- ICMP redirect events

  ICMP redirect packets and ICMPv6 redirect packets can be used to modify your routing tables. You can use IDS policy to provide notification of attempts to modify your routing tables in this manner. You can also use IDS policy to ignore ICMP redirect packets and ICMPv6 redirect packets.

  Tip: ICMP redirect packets are ignored if IPCONFIG IGNOREREDIRECT is specified in the TCP/IP profile, you are using OMPROUTE and you have IPv4 interfaces configured to OMPROUTE, or IDS policy is active for ICMP redirect attacks and the associated policy action requests that the packet be discarded. ICMPv6 redirect packets are ignored if IPCONFIG6 IGNOREREDIRECT is specified in the TCP/IP profile, you are using OMPROUTE and you have IPv6 interfaces configured to OMPROUTE, or IDS policy is active for ICMP redirect attacks and the associated policy action requests that the packet be discarded.

- Outbound raw restrictions

  Most network attacks require the ability to create packets that would not typically be built by an appropriate protocol stack implementation. You can use IDS policy to detect and prevent many of these attempts so that your system is not used as the source of attacks on other systems. As a part of this checking, you can restrict the IP protocols that are allowed in an outbound raw packet.

  Guideline: You should restrict the TCP protocol (6) on the outbound raw rule.

  Restriction: This attack type applies to IPv4 packets only. The IPv6 outbound raw restrictions attack type provides analogous function for IPv6 packets.

  You can use IDS policy to provide notification of an outbound raw packet that is considered an attack, as well as to discard the packet.

- IPv6 outbound raw restrictions

  Most network attacks require the ability to create packets that would not normally be built by an appropriate protocol stack implementation. You can use IDS policy to detect and prevent many of these attempts so that your system is not used as the source of attacks on other systems. As a part of this checking, you can restrict the protocols that are allowed in an outbound raw packet for IPv6 traffic.

  Guideline: You do not need to restrict the TCP protocol (6) for IPv6 because a TCP packet cannot be generated using AF_INET6 raw sockets.

  Restrictions:

  – You cannot restrict which IPv6 extension headers are allowed in an outbound raw packet with this attack type. The restricted protocol list applies to only the upper layer protocol for which the raw socket is opened. If any of the following extension header values are included in the restricted protocol list, it will have no effect:

    - HOPOPT (0)
    - IPv6 (41)
    - IPv6-Route (43)
    - IPv6-Frag (44)
    - ESP (50)
    - AH (51)
    - IPv6-NoNxt (59)
    - IPv6-Opts (60)

– This attack type applies to IPv6 packets only. The outbound raw restrictions attack type provides analogous function for IPv4 packets.

You can use IDS policy to provide notification of an outbound IPv6 raw packet that is considered an attack, as well as to discard the packet.

- Flood events

  Two types of floods are currently detected:

  – TCP SYN floods

    A popular denial of service attack is to flood a public server with connection requests from incorrect or nonexistent source IP addresses. The intent is to use up the available slots for connection requests and thereby deny legitimate access from completing. z/OS CS provides protection from this attack regardless of IDS policy.

  – Interface floods

    If a large number of discards are occurring in proportion to the number of inbound packets, a malicious user might be attempting a denial of service attack. If this percentage of discards for an interface exceeds a specified percentage, this is considered an interface flood. The default percentage of discards used is 10%. You can override this default by specifying the interface flood percentage parameter.

    To prevent the false detection of an interface flood condition when there is a low volume of inbound traffic on an interface, a minimum number of discards must occur in a one minute period to qualify as a flood. The default minimum is 1000 discards per minute. You can override this default by specifying the interface flood minimum discard parameter.

    When an interface flood condition is reported for an interface, the discard rate for the interface is evaluated for each subsequent 1-minute interval. An interface flood condition end is reported when the number of discards for the 1-minute interval falls below the interface flood minimum discard parameter value or the discard percentage falls below 50% of the interface flood percentage parameter value.

    If the interface flood continues for more than 5 minutes and logging was requested by policy, an interface flood continues record is logged at 5 minute intervals while the interface flood conditions exist. This log data contains additional information about the discarded packets for the interface.

    Because it can be difficult to distinguish between a malicious user trying to flood a system, unusual spikes in traffic, and problems that can be caused by setup problems, it is possible for an interface flood condition to be reported when the source of the problem is not actually a flood. For example, if enough storage is not configured to handle the inbound traffic, a large percentage of the inbound packets might be discarded and cause the interface flood percentage to be exceeded.

  This type of attack applies to both IPv4 and IPv6.

  You can use IDS policy to provide notification of an attack so that you can address the situation with your network administrators and service providers in a timely manner. Notification of a flood can include flood start and flood end event messages and tracing of the first 100 packets discarded due to the flood.

- Enterprise Extender (EE) malformed packet events

  Malformed packets for EE can be designed to disrupt the TCP/IP and SNA protocol stacks. These packets might have incorrect information in the IP and UDP headers, and the received packet might be too short or too long depending on the Logical Data Link Control (LDLC) command present. Any packet received that is not a valid EE command is flagged as malformed.

  You can use the IDS policy option to discard malformed packets or to forward them to VTAM for further processing. You can use IDS policy to provide notification of the EE malformed packets. You can code an exclusion list in the policy to exclude packets from certain IP addresses from the malformed packet checks.

  Guideline: If the existing behavior of a remote peer is flagged as an attack and you determine that it is not an actual problem, you can choose to exclude the peer from malformed packet checking by adding

the peer to the exclusion list. You can configure an exclusion by specifying the remote IP address, and optionally the remote ports.

This type of attack applies to both IPv4 and IPv6 packets.

- EE LDLC check events

  EE architecture defines five ports that map to SNA data priorities. The lowest port, usually 12000, is used for signaling data. If signaling data is received on any of the other ports, it is suspicious. If the EE LDLC check attack type is enabled, receiving signaling data on any of the other ports is detected as an attack. This type of attack can be a denial of service (DoS) attack designed to keep resources busy, so that SNA LUs are not able to start sessions with applications on this host.

  IDS policy gives you the option of discarding these packets or forwarding them to VTAM for further processing. You can configure the policy to provide notification of the LDLC events. You can code an exclusion list in the policy to exclude packets from certain IP addresses from the LDLC checks.

  Guideline: If the existing behavior of a remote peer is flagged as an attack and you determine that it is not an actual problem, you can choose to exclude the peer from LDLC event checking by adding the peer to the exclusion list. You can configure an exclusion by specifying the remote IP address, and optionally, the remote ports.

  This type of attack applies to both IPv4 and IPv6 packets.

- EE port check events

  You can use IDS policy to check the source and destination ports in the UDP header of received EE packets. The port values are expected to be the same; for example, both source and destination port are 12000 in a received packet. If the port values are not the same, this packet is considered an attack packet.

  IDS policy gives you the option of discarding these packets or forwarding them to VTAM for further processing. You can configure the policy to provide notification of the port detection events. You can code an exclusion list in the policy to exclude packets from certain IP addresses from the port checks. For example, some products send the null XID using an ephemeral source port value, but expect the reply to be sent to the EE signaling port.

  Guideline: If the existing behavior of a remote peer is flagged as an attack and you determine that it is not an actual problem, you can choose to exclude the peer from port checking by adding the peer to the exclusion list. You can configure an exclusion by specifying the remote IP address, and optionally, the remote ports.

  This type of attack applies to both IPv4 and IPv6 packets.

- EE XID flood event

  A flood of XIDs can consume all available EE lines, causing additional EE connection requests to fail. You can enable the EE XID flood attack type to receive notification when an XID flood is occurring for a specific VIPA.

  An XID flood is detected when a large number of XID exchanges time out. No discard action exists for this attack type. You can code an exclusion list in the policy to exclude XID exchanges that time out from being counted toward the XID flood if the peer's IP address is in the exclusion list.

  Guideline: If the existing behavior of a remote peer is flagged as an attack and you determine that it is not an actual problem, you can choose to exclude the peer from XID flood checking by adding the peer to the exclusion list. You can configure an exclusion by specifying the remote IP address, and optionally, the remote ports.

  This type of attack applies to both IPv4 and IPv6 packets.

- Data hiding events

  Certain fields within a packet can be used to hide data. This support enables you to detect inbound IP packets that might contain hidden data. The following checks for hidden data can be enabled:

  – Check IP option pad fields for hidden data:

- For IPv4 packets, the options field is in the IP header and can contain padding for alignment purposes.
- For IPv6 packets, a hop-by-hop options extension header or a destination options extension header can include one or more PadN options for alignment purposes.

– Check embedded packets within ICMP and ICMPv6 error messages for hidden data.

This type of attack applies to both IPv4 and IPv6 packets.

You can use IDS policy to provide notification of a suspicious packet that might contain hidden data, as well as to discard the packet.

- TCP queue size events

  You can use IDS policy to detect when the send, receive, or out-of-order queue for a TCP connection becomes constrained. A queue can be constrained due to the amount of data on the queue or due to the age of the data on the queue. When a designated amount of data that is configured in policy remains on the queue for at least 30 seconds, the queue becomes constrained. When any data remains on the queue for at least 60 seconds, the queue becomes constrained. Optionally, you can reset the TCP connection when a constraint is detected. If you do not reset the TCP connection, IDS continues to monitor the queue and detect when it becomes unconstrained. This type of attack applies to both IPv4 and IPv6 connections.

  IDS policy for TCP queue size events specifies one of four abstract queue sizes:

  – VERY_SHORT
  – SHORT
  – LONG
  – VERY_LONG

  The actual queue sizes associated with these abstract values are internal values that represent a percentage of the total size of each queue, and are subject to change. Most TCP applications have timeout values that are based on human perceptions of responsiveness. These values tend to stay constant, while system processing speeds and network delivery speeds continue to advance rapidly, which might require that the queue sizes associated with these abstract values be changed over time.

  A TCP connection might have data queued on the send queue for a long period of time for a legitimate reason. For example, a TCP connection that is associated with a printer that has run out of paper can remain in a persist state while waiting for paper to be loaded. To exclude a device, such as a printer, from TCP send queue size checking, you can configure an exclusion specifying the IP address, and optionally the port, of the device.

  You can use IDS policy to provide notification of a constrained send, receive, or out-of-order queue, as well as to reset the connection.

- Global TCP stall events

  There are attacks designed to consume system resources by creating many TCP connections and causing them to stall, making them unable to send data. This support detects a global TCP stall condition for a TCP/IP stack when at least 50% of the active TCP connections are stalled and at least 1000 TCP connections are active. Optionally, you can reset the stalled connections. This type of attack applies to both IPv4 and IPv6 connections.

For each attack category (for example, restricted IP protocol) the single highest priority rule is mapped at policy change.

One or more notification options can be specified in the action to provide the wanted documentation of detected attacks.

For IDS attack policy, the notification options enable attack events to be logged to syslogd and to the system console. The console messages provide a subset of the information provided in the syslogd messages.

For all attack categories except flood, EE XID flood, TCP queue size, and global TCP stall, a single packet triggers an event. To prevent message flooding to the system console, you should use the maximum

event message parameter to specify the maximum number of console messages to be logged per attack category within a 5-minute interval. To prevent message flooding to syslogd, a maximum of 100 event messages per attack category are logged to syslogd within a 5-minute interval. Specify the syslogd detail parameter with the global TCP stall attack type to request that a syslogd message be generated for each stalled connection when a global TCP stall is detected.

For IDS attack policy, the statistics action provides a count of the number of attack events detected during the statistics interval. The count of attacks is kept separately for each category of attack (for example, malformed packet events) and a separate statistical record is generated for each. If you want to receive statistics for attacks, you should specify exception statistics. With exception statistics, a statistical record is generated for the category of attack only if the count of attacks is nonzero. If you request normal statistics, a record is generated for every statistics interval regardless of whether an attack has been detected during that interval.

If you want to provide overrides to the interface flood parameters (interface flood minimum percentage parameter and interface flood minimum discard parameter), you should not use exception statistics. In this case, use normal statistics for a period for the flood attack category to collect data to help determine the appropriate policy parameter values. After you determine the appropriate values, specify exception statistics.

For IDS attack policy, the trace data and trace record size parameters indicate whether packets associated with attack events are to be traced. For all attack categories except flood, EE XID flood, TCP queue size, and global TCP stall, a single packet triggers an event and the packet is traced. To prevent trace flooding, a maximum of 100 attack packets per attack category are traced within a 5-minute interval.

- For the flood category, the first 100 packets discarded during a SYN flood are traced. In the case of an interface flood, the flood is detected on an interface basis and the trace limit is applied on an interface basis.
- For EE XID flood, TCP queue size, and global TCP stall, no data is traced.

For all attack categories except EE XID flood, TCP queue size, and global TCP stall, you can specify that packets associated with attack events should be discarded. However, malformed and flood packets are always discarded regardless of this setting.

For the TCP queue size attack category, you can specify that connections associated with attack events should be reset.

For the global TCP stall attack category, you can specify that stalled connections should be reset when a global TCP stall condition is detected.

The EE XID flood attack category monitors the number of XID exchange timeouts. There is no associated packet to discard or connection to reset.

An action can be unique to a specific category of attack (for example, malformed) or shared by one or more categories of attacks. If an action is shared, statistical data is still kept separately for each type of attack. Also, the maximum console message limit is enforced individually for each category of attack.

Traffic regulation policies

IDS traffic regulation (TR) policies are used to limit memory resource consumption and queue delay time during peak loads.

Traffic regulation policies for TCP ports

IDS traffic regulation (TR) policies for TCP ports limit the total number of connections an application has active at one time. This can be used to limit the number of address spaces created by forking applications such as FTPD and otelnetd. A *fair share* algorithm is also provided based on the percentage of remaining available connections already held by a source IP address.

The percentage is applied against the number of available connections for the port. Therefore, as fewer connections become available, each host is allowed fewer new connections. The percentage is applied

against the number of available connections, rather than the total number of connections allowed, in order to allow access to a larger number of different hosts when resources are low.

When a host requests a connection, the number of connections it currently holds for the port is compared to the percentage applied to the connections currently available for the port. If the number currently held is less than the percentage of currently available connections, the host is allowed to open an additional connection. If equal or greater, the host is not allowed to open further connections until more connections are freed up. All connection requesters for the port are regulated by this mechanism. If a host does not currently have any connections open on the port and unused connections are available, a host will always be allowed at least 1 connection. Multi-user source IP addresses can be allowed a larger number of connections by specifying a QoS policy with a higher number of connections (MaxConnections) than allowed by the TR policy. TR will honor the QoS differentiated services policy if the port is not in a *constrained state*. A QoS exception is made only when QoS differentiated service policy is applied for the specific source server port and specific outbound client destination IP address; if either of these attributes specify a range or are null, the QoS exception will not be made.

TR TCP generates a Constrained Event when a port reaches about 90% of its Connection Limit. An Unconstrained Event is generated when the port falls below about 88% of its limit. An IDS correlator is assigned for the duration of each constrained state. If tracing is requested in the policy, the first 100 packets that exceed the limit in each constrained state are traced along with the correlator. TR TCP also generates events for each connection allowed because of a QoS override policy and for each connection denied for exceeding either the application's connection limit or the percent available limit.

To prevent possible flooding of syslog, TR TCP limits the number of connection refused, would have been refused, or QOS exception log records written in a five minute interval. For a listening port, a maximum of 100 of these log records are written within a five minute interval. Globally, TR TCP writes a maximum of 1000 of these log records within a five minute interval. If a log record was not written due to these limits, the count of refused or would have been refused connection log records that were not logged is recorded in the `EZZ8660I TRMD TCP connection log records suppressed` log message after the five minute interval ends. Similarly, the count of QOS exception records that were not written is recorded in the `EZZ8661I TRMD TCP QOS exception log records suppressed` log message.

TCP traffic regulation policy applies to IPv4 and IPv6.

Guideline: TR TCP is intended for use with long-running servers, which typically use well-known ports or ports that are reserved for TCP applications. You should not use TR TCP to monitor transitory listeners; doing so can result in high storage usage. An example of a transitory listener is an FTP data connection that lasts only as long as it takes to move one file and typically uses ports above 1023.

Traffic regulation policies for UDP ports

Previously, control over UDP based applications consisted of application priority management and the TCP/IP profile parameter UDPQueueLimit ON | OFF. Inbound datagrams for bound UDP ports are accepted and queued until the queue limit is reached or buffer memory is exhausted. If UDPQueueLimit is set to OFF, any single bound port under a flood attack or with a stalled application could consume all available buffer storage. It is recommended that UDPQueueLimit always be set to ON. This limits the amount of storage that can be consumed by inbound datagrams for any single bound port. Sockets that use the Pascal API, have a limit of 160 KB in any number of datagrams. Sockets that use other APIs, have a limit of 2000 datagrams or 2880 KB.

IDS traffic regulation (TR) policies for UDP ports specify one of four abstract queue sizes for specified bound IP addresses and ports. The four abstract sizes are VERY_SHORT, SHORT, LONG and VERY_LONG. The actual queue sizes associated with these abstract values are internal values that are subject to change. Most UDP applications have timeout values based on human perceptions of responsiveness. These values tend to stay constant while system processing speeds and network delivery speeds continue to advance rapidly. This might require the queue sizes associated with these abstract values to change over time. The initial implementation uses the values of 16, 256, 2048 and 8192 for the number of datagrams and an average datagram size of 2 KB to calculate the byte sizes (32 KB, 512 KB, 4 MB, 16 MB). For performance reasons, sockets that use the Pascal API will enforce only the byte limit. Sockets that use other APIs will enforce both limits. Sockets without a policy specified for their port will use the existing UDPQueueLimit mechanism.

For applications that can process datagrams at a rate faster than the average arrival rate, the queue acts as a speed matching buffer that shifts temporary peak workloads into following valleys. The more that the application processing rate exceeds the average arrival rate and the larger the queue, the greater the variation in arrival rates that can be absorbed without losing work. Very fast applications with highly variable ("bursty") traffic patterns might benefit from LONG or VERY_LONG queue sizes.

For applications that consistently receive datagrams at a higher rate than they are able to process them, the queue limits the effective arrival rate to the processing rate by discarding excess datagrams. In this case the queue size influences only the average wait time of datagrams in the queue and not the percentage of work lost. In fact, if the wait time gets too large, the peer application might have given up or retransmitted the datagram before it is processed. Slow applications with consistently high traffic rates might benefit from SHORT queue sizes.

In general, client side applications will tend to have lower system priority giving them lower datagram processing rates. They also tend to have much lower datagram arrival rates. Giving them SHORT or VERY_SHORT queue sizes might reduce the risk to system buffer storage under random port flood attacks with little impact on percentage of datagrams lost.

TR UDP generates a Constrained Event when a port reaches about 90% of its Queue Limit. An Unconstrained Event is generated when the port falls below about 88% of its limit. An IDS correlator is assigned for the duration of each constrained state. If tracing is requested in the policy, the first 100 packets that exceed the limit in each constrained state are traced along with the correlator.

UDP traffic regulation policy applies to IPv4 and IPv6.

Options for configuring IDS

You configure IDS using a set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the IDS policy for each TCP/IP stack. In a complex environment, this file can become large. For this reason, there are two alternatives for creating the Policy Agent files.

Option 1: Use the IBM Configuration Assistant for z/OS Communications Server

The IBM Configuration Assistant for z/OS Communications Server, a GUI-based tool, provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the Policy Agent files.

The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

Through a series of wizards and online help panels, you can use the IBM Configuration Assistant for z/OS Communications Server to create IDS configuration files for any number of z/OS images with any number of TCP/IP stacks per image. Using the IBM Configuration Assistant for z/OS Communications Server, there are two types of reusable objects:

- Traffic descriptors that define the IP traffic type, such as TCP or UDP.
- Requirement maps that contain attack protection, scan protection, and traffic regulation. For scan protection and traffic regulation, traffic descriptors are used to identify the local applications that are provided the protection and regulation. A single requirement map should contain a complete set of IDS requirements that will govern the level of IDS for a TCP/IP stack.

For each TCP/IP stack, you select a requirement map that provides IDS for the stack. The IBM Configuration Assistant for z/OS Communications Server comes with a number of IBM-supplied traffic descriptors and a default requirement map that are easily applied, or you can use the IBM-supplied definitions as the basis for your own set of reusable objects.

The IBM Configuration Assistant for z/OS Communications Server can dramatically reduce the amount of time that is required to create IDS policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS security configuration, using the GUI can help you ensure that you have a consistent and easily manageable interface for implementing IDS.

The information in "Defining IDS policies" on page 864 primarily describes option 2, manual configuration. However, if you are using the IBM Configuration Assistant for z/OS Communications Server, reading this information will help you understand security concepts and the relationship between Policy Agent and IDS function.

Option 2: Manual configuration

You can manually create the IDS policy configuration files by coding all of the required statements in a file. There are a large number of configuration options provided by IDS policy statements that permit advanced users to carefully fine-tune IDS policy on a per-stack basis. The information in "Defining IDS policies" on page 864 describes the procedure for creating an IDS policy by manually creating and editing the configuration files. For details about the IDS policy statements, see z/OS Communications Server: IP Configuration Reference.

Specifying the IDS configuration file based on Policy Agent role

The Policy Agent can act as a policy server, a policy client, or neither. For more information on these different roles, see "Policy types and infrastructure overview" on page 787. Regardless of which option is used to configure IDS policies, the resulting configuration files need to be specified using different statements, depending on the role of the Policy Agent.

- If you are using the Policy Agent as a policy client that retrieves IDS policies from the policy server, specify the configuration files using the DynamicConfigPolicyLoad statement on the policy server.
- If you are using the Policy Agent as a policy client, but the policy client does not retrieve IDS policies from the policy server, specify the configuration files using the IDSConfig statement on the policy client.
- If you are not using a policy client/policy server environment, specify the configuration files using the IDSConfig statement on the single Policy Agent.

When this information refers to configuration files, keep in mind where the files should exist, based on the role of the Policy Agent.

Defining IDS policies

IDS policies are stored in a Policy Agent IDS configuration file, a server that supports LDAP, or both. IDS policies are processed by Policy Agent and installed into a z/OS Communications Server TCP/IP stack. Before creating IDS policies, you should be familiar with the information about running Policy Agent, the IDS configuration file, and LDAP in Chapter 14, "Policy-based networking," on page 787.

Restriction: Not all IDS policy options are available in an LDAP configuration file. Each of the policy example sections, IDS scan policy example, IDS attack policy examples, and traffic regulation policy example, indicate which policy options are not available if you are using LDAP to define IDS policy.

A conservative approach to defining IDS policy will avoid unexpected application outages and excessive rule processing. The examples here describe policies provided in the sample files shipped with the system. (See "Policy sample files" on page 800.)

IDS policy definition considerations

IDS policies can be defined with different condition type values.

Tip: Condition type is configured in the IDS configuration file with the ConditionType statement, and in the LDAP configuration with the ibm-idsConditionType attribute.

The supported condition types are as follows:

SCANGLOBAL
This policy is searched by condition type only. The highest priority scan global rule is mapped at policy change and cached. The policy defines the FastScan and SlowScan parameters, and the reporting and tracing actions to take when a scan is detected. The statistics reporting option is not applicable to scan processing. Scan processing does not impose limits, discard packets, or reset connections.

SCANEVENT
These policies are searched by condition type and a protocol condition of ICMP, ICMPv6, TCP, or UDP. For protocols TCP and UDP, the policy search also includes the local destination port and the bound IP address. For ICMP and ICMPv6, the highest priority scan event rule is mapped at policy change and cached. The TCP and UDP rules are mapped when a potentially countable event occurs. If the event is associated with a bound socket, the rule is cached. The policies associated with these rules define the sensitivity level to use for counting events towards the scan thresholds and the source exclusion list to use for the mapped events. Packet tracing occurs if the action associated with the scan global rule activates tracing and the sensitivity level indicates that the event is countable. The statistics reporting option is not applicable to scan processing. Scan processing does not impose limits, discard packets, or reset connections.

ATTACK
There are several attack types. For each attack type, the highest priority rule is mapped at policy change and cached. The reporting and statistics actions are supported for all attack types. The tracing action is supported for all attack types except EE XID flood, TCP queue size, and global TCP stall. For each attack type, the policy is searched by only condition type and attack type.

Tip: You can configure a discard action in the IDS configuration file as ActionType ATTACK DISCARD, and in the LDAP configuration using the ibm-idsTypeActions:LIMIT attribute.

Other supported actions are defined for each attack type. The supported attack types are:

MALFORMED_PACKET
Malformed packets are always discarded by the stack, even if no discard was requested by the policy.

FLOOD
Flood packets are always discarded by the stack, even if no discard was requested by the policy.

ICMP_REDIRECT
ICMP redirect packets are discarded if this policy specifies discard.

IP_FRAGMENT
If this policy specifies discard, fragmented datagrams are discarded if a fragmentation overlay is detected that changes the data in the packet, including changes to the length of the packet.

RESTRICTED_IP_OPTIONS
A list of restricted IP options is specified for this attack type. IP option 0 (end of list) and 1 (NO-OP) cannot be restricted and are ignored if specified. If this policy specifies discard, IPv4 packets are discarded if they contain an IP option that is specified in the list of restricted IP options.

RESTRICTED_IPV6_DST_OPTIONS
A list of restricted IPv6 destination options is required for this attack type. You cannot restrict options 0 (Pad1) or 1 (PadN); they are always allowed. If this policy specifies discard, IPv6 packets are discarded if they contain an IPv6 destination options extension header with an option that is specified in the list of restricted IPv6 destination options.

RESTRICTED_IPV6_HOP_OPTIONS
A list of restricted IPv6 hop-by-hop options is required for this attack type. You cannot restrict options 0 (Pad1) or 1 (PadN); they are always allowed. If this policy specifies discard, IPv6 packets are discarded if they contain an IPv6 hop-by-hop extension header with an option that is specified in the list of restricted IPv6 hop-by-hop options.

RESTRICTED_IP_PROTOCOL
A list of restricted IP protocols is required for this attack type. IP protocols 1 (ICMP), 6 (TCP), and 17 (UDP) cannot be restricted and are ignored if specified. If this policy specifies discard,

IPv4 packets are discarded if they contain a protocol that is specified in the list of restricted IP protocols.

RESTRICTED_IPV6_NEXT_HDR
A list of restricted IPv6 next header values is required for this attack type. You cannot restrict next header values 6 (TCP), 17 (UDP), or 58 (ICMPv6); they are always allowed. If this policy specifies discard, IPv6 packets are discarded if they contain a next header value that is specified in the list of restricted IPv6 next headers.

OUTBOUND_RAW
A list of restricted IP protocols is required for this attack type. If this policy specifies discard, outbound IPv4 raw packets that meet any of the following criteria are discarded:

- Written to a raw socket that has a source IP address that is not in the stack's home list
- Fragmented by the application
- Specifies one of the ICMP reply types
- Specifies a protocol that is in the list of restricted IP protocols

OUTBOUND_RAW_IPV6
A list of restricted IP protocols is required for this attack type. If this policy specifies discard, outbound IPv6 raw packets that meet any of the following criteria are discarded:

- Specifies one of the ICMPv6 reply types
- Specifies one of the ICMPv6 neighbor discovery (ND) types
- Specifies one of the ICMPv6 multicast listener discovery (MLD) types
- Specifies a protocol that is in the list of restricted IP protocols

PERPETUAL_ECHO
A list of local UDP ports and a list of remote UDP ports is necessary with this attack type. Each port list is limited by the stack to the first 20 ports specified. The ports in each inbound UDP packet are checked against these lists. The destination port is checked against the local port list. The source port is checked against the local port list if the source IP address is in the stack's home list. Otherwise, the source port is checked against the remote port list. If this policy specifies discard, UDP packets with both ports in the checked port lists are discarded.

When defining the policy in LDAP, this attack type condition must be specified in a complex rule using CNF and multiple condition levels. The attack type condition is at one of the condition levels. There must be a list of conditions defining the local port list at a second level. There must be a list of conditions defining the remote port list at a third level. The negated flag is ignored by the stack on port list conditions.

EE_LDLC_CHECK
EE signaling data received on a port other than the signaling port is discarded if this policy specifies discard. An exclusion list is applied if one is configured. An EE packet is not flagged as an attack if the source IP address is found in the exclusion list.

EE_PORT_CHECK
EE packets with different source and destination ports are discarded if this policy specifies discard. An exclusion list is applied if one is configured. An EE packet is not flagged as an attack if the source IP address is found in the exclusion list.

EE_MALFORMED_PACKET
Malformed EE packets are discarded if this policy specifies discard. An exclusion list is applied if one is configured. A malformed EE packet is not flagged as an attack if the source IP address is found in the exclusion list.

EE_XID_FLOOD
An EE XID flood is detected when a large number of EE XID exchanges time out. No discard action is associated with this attack type. An exclusion list is applied if one is configured. An EE XID time out is not counted as part of flood detection if the source address of the XID is found in the exclusion list.

DATA_HIDING
If this policy specifies discard and enables checking of IP option pad fields or embedded packets within an ICMP error message, packets containing hidden data are discarded.

TCP_QUEUE_SIZE
If this policy specifies reset, the TCP connection is reset if the send, receive, or out-of-order queue becomes constrained. A queue can be constrained due to the amount of data on the queue or the age of the data on the queue. A queue size is specified in the conditions. A list of remote IP addresses, that are to be excluded when monitoring the send queue, can optionally be specified in the conditions.

GLOBAL_TCP_STALL
If this policy specifies reset, the stalled TCP connections are reset when a global TCP stall condition is detected.

TR

These policies are searched by condition type, protocol (TCP or UDP), local IP address, and local port. TCP rules are mapped when a local application does a listen on a socket or when an inbound connection handshake is completed. UDP rules are mapped when an inbound packet arrives at a local bound socket. UDP TR policy supersedes the TCPIP PROFILE setting of UDPQUEUELIMIT for covered ports. Mapped rules are cached and associated with the bound socket.

For TCP, the policy defines the total number of allowed connections, the percentage of remaining available connections any single source IP can acquire and whether these limits are applied globally across all applications using this port number or applied individually to each application using an instance of this port number. For UDP, the policy defines which of the four available queue sizes is applied to each application using this port number. TR actions define the reporting, statistics and tracing actions for covered ports. If the policy specifies action LIMIT, connections or packets that exceed the limits are discarded.

Notes:

1. For TCP, a total connection limit or percentage available limit of zero, with an action of LIMIT effectively quiesces the application.
2. For TCP, a local host IP address cannot be specified in any condition if a TR TCP limit scope value of PORT is specified.
3. For UDP, a policy for a port without an action of LIMIT effectively makes the application unlimited.
4. Each LDAP IDS TR action must specify at least one ibm-idsTypeActions attribute.

IDS scan policy example

The goal of scan policy is to detect all scanners with potentially malicious intent while avoiding large numbers of false positives. You can make this process more efficient by reserving all unused low ports in the TCPIP profile. This will allow you to use the low sensitivity setting on scans for these ports. As you investigate the scans detected, you will initially find your own network management tools. These can be explicitly excluded. If you include UDP ephemeral ports in a high sensitivity policy, you will discover that your DNS servers show up as scanners. You can explicitly exclude these as well. To activate scan policy, a scan global rule and at least one scan event rule must be defined.

The following scan rules are defined:

- Scan Global

  Defines a global set of parameters for detecting scans, and also defines reporting parameters for scan events.

  – A Fast Scan is defined as 5 unique events in 2 minutes from a single source IP address.
  – A Slow Scan is defined as 10 unique events in 480 minutes (8 hours).
  – The first 200 bytes of the packet associated with each countable event will be traced.
  – When a scan is detected an event will be written to syslog warning level, along with a detailed list of all the unique events included in the scan.

  - No message will be written to the console.
- Scan Event Low

  Defines a set of traffic for which low sensitivity scan detection will be performed. Inbound traffic to all TCP and UDP ports between 1 and 1023 will be monitored. It is recommended that unused low ports be RESERVED in the TCPIP Profile.
- Scan Event Medium

  Defines a set of traffic for which medium sensitivity scan detection will be performed. ICMP and ICMPv6 inbound traffic will be monitored.

The following example is an IDS configuration file:

```
##################################################################################
##################################################################################
# Scan Policies
##################################################################################

#---------------------------------------------------------------------
# Scan - IDSRule
#---------------------------------------------------------------------
IDSRule                     ScanEventLowTcp-rule
{
  ConditionType             ScanEvent
  Priority                  2
  IDSScanEventConditionRef ScanTcpLowCondition
  IDSActionRef              ScanEventLow-action
}
IDSRule                     ScanEventLowUdp-rule
{
  ConditionType ScanEvent
  Priority                  2
  IDSScanEventConditionRef ScanUdpLowCondition
  IDSActionRef              ScanEventLow-action
}
IDSRule                     ScanEventMedium-rule
{
  ConditionType             ScanEvent
  Priority                  2
  IDSScanEventCondition
  {
    Protocol                Icmp
  }
  IDSActionRef              ScanEventMedium-action
}
IDSRule                     ScanEventMedium-rule-v6
{
  ConditionType             ScanEvent
  Priority                  2
  IDSScanEventCondition
  {
    Protocol                Icmpv6
  }
  IDSActionRef              ScanEventMedium-action
}
IDSRule                     ScanGlobal-rule
{
  ConditionType             ScanGlobal
  Priority                  2
  IDSActionRef              ScanGlobal-action
  IDSScanGlobalCondition                  # inline condition
  {
    FSinterval              2
    SSInterval              480
  }
}

#---------------------------------------------------------------------
# Scan - IDSScanEventCondition
#---------------------------------------------------------------------
IDSScanEventCondition      ScanTcpLowCondition
{
  Sensitivity               Low
  Protocol                  Tcp
  LocalPortRange            1 1023
}
IDSScanEventCondition      ScanUdpLowCondition
```

```
{
  Sensitivity              Low
  Protocol                 17              # Udp
  LocalPortRange           1 1023
}

#---------------------------------------------------------------------
# Scan - IDSAction
#---------------------------------------------------------------------
IDSAction                  ScanEventLow-action
{
  ActionType               ScanEvent count
}
IDSAction                  ScanEventMedium-action
{
  ActionType               ScanEvent count
}
IDSAction                  ScanGlobal-action
{
  ActionType               ScanGlobal
  IDSReportSet             ScanGlobalReportSet
  {
    TypeActions            Log
    LogDetail              Yes
    TraceData              RecordSize
    TraceRecordSize        200
  }
}
```

If you are using LDAP to define policy, see "IDS scan policy example" on page 1397.

Restrictions: LDAP policy cannot be used to:

- Define that ICMPv6 traffic should be monitored for scan events
- Exclude IPv6 addresses in the scan exclusion list

IDS attack policy examples

The goal of attack policy is to help protect your system from both known and unknown attacks and to give you timely notification when attacks do occur.

A malformed packet attack policy covers many known attacks designed to cause system failures. These packets are always discarded and rarely have legitimate source address information. Many malformed packet attacks use fragmentation to overlay header fields. The IDS fragment restriction policy is intended to protect you from unknown attacks of this type by detecting fragment overlays that change the data in a packet, including changes to the length of the packet.

Unless you know that you need to process ICMP and ICMPv6 redirect messages, you should disallow them with policy.

There are several types of flood attacks. IDS can identify TCP SYN floods and interface floods. IDS policy should be used to notify you when a flood occurs. You will need to work with your network administrators and service providers to track the flood backwards, one physical hop at a time, to locate the sources.

The IP protocol restrictions and IP option restrictions provide additional protection against future unknown attacks. These are specific to IPv4. The philosophy behind them is to disallow anything that you do not have a known reason to allow. Similar protection is provided for IPv6 by the IPv6 next header restrictions, IPv6 destination option restrictions, and IPv6 hop-by-hop option restrictions.

A data hiding attack policy provides notification when an inbound packet might contain hidden data.

The global TCP stall detection provides protection against an attack that consumes systems resources by causing many TCP connections to stall, unable to send data.

The outbound raw policy is intended to help you detect someone using your system as the base for an attack. It looks for several behaviors associated with *spoofed* packets. This protection applies to IPv4 sockets. The IPv6 outbound raw policy provides similar protection for IPv6 sockets.

Attack rules define the set of conditions that define what constitutes an attack for a given attack type. The highest priority rule of each attack type is used. The action associated with an attack rule defines reporting and logging options for a detected attack.

The following types of attack rules are defined in the sample policy:

- Malformed packet

  Protects against various types of known attacks based on malformed packets. Applies to IPv4 and IPv6 packets.

- Flood

  Protects against TCP SYN flood and interface flood attacks. Applies to IPv4 and IPv6 connections and interfaces.

- ICMP redirect

  Ignores received ICMP and ICMPv6 redirect messages.

- IP fragment

  Disallows fragmentation overlays that change the data in a packet, including changes to the length of the packet. Applies to both IPv4 and IPv6 packets.

- IP protocol restrictions

  Defines restricted IP protocols, and restricts everything except ICMP, TCP, and UDP. Applies to IPv4 packets only.

- IPv6 next header restrictions

  Defines restricted IPv6 next header values, and restricts everything except the following values:

  – 0 (Hop-by-hop options extension header)
  – 6 (TCP protocol)
  – 17 (UDP protocol)
  – 41 (IPv6 extension header)
  – 43 (routing extension header)
  – 44 (fragmentation extension header)
  – 50 (ESP extension header)
  – 51 (AH extension header)
  – 58 (ICMPv6 protocol)
  – 59 (No next header)
  – 60 (Destination options extension header)
  – 89 (OSPF protocol)
  – 135 (Mobility header)

  Applies to IPv6 packets only.

- IPv6 destination option restrictions

  Defines restricted options in the IPv6 destination extension header, and restricts all options except the following options:

  – 0 (Pad1)
  – 1 (PadN)
  – 4 (Tunnel encapsulation limit)
  – 5 (Router alert)
  – 6 (Quick-start)
  – 7 (CALIPSO)
  – 138 (Endpoint identification)

  - 194 (Jumbo payload)
  - 201 (Home address)

  Applies to IPv6 packets only.
- IPv6 hop-by-hop option restrictions

  Defines restricted options in the IPv6 hop-by-hop extension header. A common IpOptionGroup is referenced by the IPv6 destination and IPv6 hop-by-hop option rules. Applies to IPv6 packets only.
- Outbound raw restrictions

  Validity checking for outbound packets using raw sockets. Restricts everything except ICMP, UDP, IGMP, and OSPFIGP. Applies to IPv4 packets only.
- EE LDLC check

  Validity checking of the Logical Data Link Control (LDLC) command for inbound EE packets to ensure that the packet is received on the correct port. A packet with a source IP address in the exclusion list is not checked. Applies to IPv4 and IPv6 packets.
- IPv6 outbound raw restrictions

  Validity checking for outbound packets using IPv6 raw sockets. Restricts everything except ICMPv6, UDP, and OSPF protocols. Applies to IPv6 packets only.
- Data hiding checks

  Applies to IPv4 and IPv6 packets.
- TCP queue size restrictions

  The queue size is defined as SHORT. Applies to IPv4 and IPv6 connections.
- Global TCP stall detection

  Applies to IPv4 and IPv6 connections.
- A single reusable attack action is defined and shared among all the attack rules and specifies the following actions:
  - Events are written to syslog ALERT level.
  - Events are not written to the system console.
  - The first 200 bytes of packets associated with an attack are traced.
  - Statistics are evaluated every 60 minutes and written only if an attack occurred.
  - In the IDS configuration file, `Nodiscard` was specified, so packets that are associated with the following attack types are not discarded:
    - ICMP redirect
    - IP fragment
    - IP protocol restrictions
    - IPv6 next header restrictions
    - IPv6 destination options
    - IPv6 hop-by-hop options
    - Outbound raw
    - IPv6 outbound raw
    - Data hiding
    - EE LDLC check

    In the LDAP example in “IDS attack policy example” on page 1399, `Limit` was not specified, which has the same effect as `Nodiscard`.
  - In the IDS configuration file, `NoResetConn` was specified, so when a TCP queue size or global TCP stall condition is detected, the TCP connections are not reset.

The following example is an IDS configuration file:

```
###############################################################################
###############################################################################
# Attack Policies
###############################################################################

#---------------------------------------------------------------------
# Attack - IDSRule
#---------------------------------------------------------------------
IDSRule                     AttackMalformed-rule
{
  ConditionType             Attack
  Priority                  2
   IDSAttackCondition
  {
    AttackType              MALFORMED_PACKET
  }
  IDSActionRef              Attack-action
}
IDSRule                     AttackFlood-rule
{
  ConditionType             Attack
  Priority                  2
  IDSAttackCondition        AttackFloodCondition
  {
    AttackType              FLOOD
    IfcFloodPercentage      10
    IfcFloodMinDiscard      1000
  }
  IDSActionRef              Attack-action
}
IDSRule                     AttackICMPRedirect-rule
{
  ConditionType             Attack
  Priority                  2
  IDSAttackCondition
  {
    AttackType              ICMP_REDIRECT
  }
  IDSActionRef              Attack-action
}
IDSRule                     AttackIpFragment-rule
{
  ConditionType             Attack
  Priority                  2
  IDSAttackCondition
  {
    AttackType              IP_FRAGMENT
  }
  IDSActionRef              Attack-action
}
IDSRule                     AttackIPProt-rule
{
  ConditionType             Attack
  Priority                  2
  IDSAttackCondition
  {
    AttackType                RESTRICTED_IP_PROTOCOL
    ProtocolGroupRef          IpProtRestrictedGroup
  }
  IDSActionRef              Attack-action
}
IDSRule                    AttackIPv6NextHeader-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             RESTRICTED_IPV6_NEXT_HDR
    IPv6NextHdrGroupRef    IPv6NextHdrGroup
  }
  IDSActionRef             Attack-action
}

IDSRule                    AttackIPv6DstOptions-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
```

```
    AttackType             RESTRICTED_IPV6_DST_OPTIONS
    RestrictedIpv6OptionGroupRef Ipv6OptGroup
  }
  IDSActionRef             Attack-action
}

IDSRule                    AttackIPv6HopByHopOptions-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             RESTRICTED_IPV6_HOP_OPTIONS
    RestrictedIpv6OptionGroupRef Ipv6OptGroup
  }
  IDSActionRef             Attack-action
}
IDSRule                     AttackOutboundRaw-rule
{
  ConditionType             Attack
  Priority                  2
  IDSAttackCondition
  {
    AttackType              OUTBOUND_RAW
    ProtocolGroupRef     IpProtOutboundRawGroup
  }
  IDSActionRef              Attack-action
}
IDSRule                    AttackIPv6OutboundRaw-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             OUTBOUND_RAW_IPv6
    ProtocolGroupRef       IpProtOutboundRaw6Group
  }
  IDSActionRef             Attack-action
}

IDSRule                    AttackDataHiding-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             DATA_HIDING
    OptionPadChk           Enable
    IcmpEmbedPktChk        Enable
  }
  IDSActionRef             Attack-action
}

IDSRule                    AttackTcpQueueSize-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             TCP_QUEUE_SIZE
    TcpQueueSize           Short
  }
  IDSActionRef             Attack-action
}

IDSRule                    AttackTcpStall-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             GLOBAL_TCP_STALL
  }
  IDSActionRef             Attack-action
}
IDSRule                    EELDLCCheck-rule
{
  ConditionType            Attack
  Priority                 2
  IDSAttackCondition
  {
    AttackType             EE_LDLC_Check
```

```
    IDSExclusion
    {
      ExcludedAddrPort    1.1.1.1
      ExcludedAddrPort    2.0.0.0/8
    }
  }
  IDSActionRef            Attack-action
}
#---------------------------------------------------------------------
# Attack - IDSAction
#---------------------------------------------------------------------
IDSAction                  Attack-action
{
  ActionType               Attack nodiscard
  ActionType               Attack noresetconn
  IDSReportSetRef          LogExceptStatReportSet
}


#---------------------------------------------------------------------
# IDSReportSet
#---------------------------------------------------------------------
IDSReportSet             LogExceptStatReportSet
{
  TypeActions            Log
  TypeActions            Statistics
  LoggingLevel           1
  StatType               Exception
  TraceData              RecordSize
  TraceRecordSize        200
}

#---------------------------------------------------------------------
# IPProtocol
#---------------------------------------------------------------------
IpProtocolRange            IpProt2to5
{
    IpProtocol             2 5
}
IpProtocolRange            IpProt7to16
{
    IpProtocol             7 16
}
IpProtocolRange            IpProt18to255
{
    IpProtocol             18 255
}
IpProtocolRange            IpProt3to16
{
  IpProtocol               3 16
}
IpProtocolRange            IpProt18to88
{
  IpProtocol               18 88
}
IpProtocolRange            IpProt90to255
{
  IpProtocol               90 255
}
IpProtocolRange            IpProt0to16
{
  IpProtocol               0 16
}
IpProtocolRange            IpProt18to57
{
  IpProtocol               18 57
}
IpProtocolRange            IpProt59to88
{
  IpProtocol               59 88
}
IpProtocolGroup            IpProtRestrictedGroup
{
  IpProtocolRangeRef       IpProt2to5
  IpProtocolRangeRef       IpProt7to16
  IpProtocolRangeRef       IpProt18to255
}
IpProtocolGroup            IpProtOutboundRawGroup
{
  IpProtocolRangeRef       IpProt3to16
  IpProtocolRangeRef       IpProt18to88
  IpProtocolRangeRef       IpProt90to255
```

```
}
IpProtocolGroup         IpProtOutboundRaw6Group
{
  IpProtocolRangeRef    IpProt0to16
  IpProtocolRangeRef    IpProt18to57
  IpProtocolRangeRef    IpProt59to88
  IpProtocolRangeRef    IpProt90to255
}
#-------------------------------------------------------------------------
# IPv6NextHeader
#-------------------------------------------------------------------------
IPv6NextHdrRange     IPv6NextHdr1to5
{
  IPv6NextHdr        1 5
}
IPv6NextHdrRange     IPv6NextHdr7to16
{
  IPv6NextHdr        7 16
}
IPv6NextHdrRange     IPv6NextHdr18to40
{
  IPv6NextHdr        18 40
}
IPv6NextHdrRange     IPv6NextHdr42
{
  IPv6NextHdr        42 42
}
IPv6NextHdrRange     IPv6NextHdr45to49
{
  IPv6NextHdr        45 49
}
IPv6NextHdrRange     IPv6NextHdr52to57
{
  IPv6NextHdr        52 57
}
IPv6NextHdrRange     IPv6NextHdr61to88
{
  IPv6NextHdr        61 88
}
IPv6NextHdrRange     IPv6NextHdr90to134
{
  IPv6NextHdr        90 134
}
IPv6NextHdrRange     IPv6NextHdr136to255
{
  IPv6NextHdr        136 255
}
IPv6NextHdrGroup     IPv6NextHeaderGroup
{
  IPv6NextHdrRangeRef IPv6NextHdr1to5
  IPv6NextHdrRangeRef IPv6NextHdr7to16
  IPv6NextHdrRangeRef IPv6NextHdr18to40
  IPv6NextHdrRangeRef IPv6NextHdr42
  IPv6NextHdrRangeRef IPv6NextHdr45to49
  IPv6NextHdrRangeRef IPv6NextHdr52to57
  IPv6NextHdrRangeRef IPv6NextHdr61to88
  IPv6NextHdrRangeRef IPv6NextHdr90to134
  IPv6NextHdrRangeRef IPv6NextHdr136to255
}
#-------------------------------------------------------------------------
# IPv6Option
#-------------------------------------------------------------------------
IpOptionGroup           Ipv6OptGroup
{
  IpOptionRangeRef      IpOpt2to3
  IpOptionRangeRef      IpOpt8to137
  IpOptionRangeRef      IpOpt139to193
  IpOptionRangeRef      IpOpt195to200
  IpOptionRangeRef      IpOpt202to255
}
IpOptionRange        IpOpt2to3
{
  IpOption           2 3
}
IpOptionRange        IpOpt8to137
{
  IpOption           8 137
}
IpOptionRange        IpOpt139to193
{
  IpOption           139 193
}
```

```
IpOptionRange        IpOpt195to200
{
  IpOption           195 200
}
IpOptionRange        IpOpt202to255
{
  IpOption           202 255
}
```

If you are using LDAP to define policy, see "IDS attack policy example" on page 1399.

Restriction: LDAP policy cannot be used to define rules for the following attack types:

- DATA_HIDING
- OUTBOUND_RAW_IPV6
- RESTRICTED_IPV6_DST_OPTIONS
- RESTRICTED_IPV6_HOP_OPTIONS
- RESTRICTED_IPV6_NEXT_HDR
- TCP_QUEUE_SIZE
- GLOBAL_TCP_STALL
- EE_MALFORMED_PACKET
- EE_LDLC_CHECK
- EE_PORT_CHECK
- EE_XID_FLOOD

Traffic Regulation policy examples

The goal of Traffic Regulation (TR) policy is to protect your system from usage spikes. A phased approach to determine the correct policy for your system is recommended.

To gather baseline statistics, you should first run in normal statistics mode, with the traffic regulation management daemon (TRMD) running. In normal statistics mode, the following information is provided for the port on a policy defined interval:

- Total number of connections requested during the interval
- Total number of connections closed during the interval
- The IP address of the host that requested a connection during the interval and held the highest number of concurrent connections during the interval, and the highest number of concurrent connections held by this IP address
- A suggested value for TotalConnections based on this interval
- A suggested value for Percentage based on this interval

While the baseline statistics records provide suggested policy values for the interval, you should evaluate data from multiple intervals. The values suggested are those that would avoid denying any of the connections in the interval. Choose lower values if the interval represents a workload larger than you want to allow.

After you determine the policy values to use, try running with the Log and Nolimit actions specified. Specifying the Nolimit action basically tests out the policy. The connections that would have been denied (if the Limit action was specified) are logged, but the connection is allowed to occur. After you are satisfied with the experimental policy, the policy action can be set to Limit.

The following traffic regulation TCP rules are defined:

- TRTcp-rule: Defines TCP baseline statistics gathering for the low port range.

  This temporary rule provides statistical reports to determine normal traffic patterns for several applications. After the baseline values are determined, this rule should be replaced by rules that include the specific conditions to be regulated.

- TRTcpWeb-rule: Defines application limits and host percentage limits for a single application.
  - This rule enforces set limits.
  - The rule has a higher priority than the TR TCP rule.
  - The rule is limited to a single server application that is bound to a specific IP address.

The following example is an IDS configuration file:

```
###########################################################################
###########################################################################
# TR Policies
###########################################################################

#---------------------------------------------------------------------
# TR - IDSRule
#---------------------------------------------------------------------
IDSRule                     TRTcpWeb-rule
{
  ConditionType             TR
  Priority                   7
  IDSTRConditionRef         TRTcpWebCondition
  IDSActionRef              TRTcpLimit-action
}
IDSRule                     TRTcp-rule
{
  ConditionType             TR
  Priority                  2
  IDSTRConditionRef         TRTcpCondition
  IDSActionRef              TRTcpLog-action
  IpTimeConditionRef        Time1
}

#---------------------------------------------------------------------
# TR - IDSTRCondition
#---------------------------------------------------------------------
IDSTRCondition              TRTcpWebCondition
{
  Protocol                  Tcp
  LocalPortRange            80
  LocalHostAddr             10.14.243.87
  TRtcpTotalConnections     1000
  TRtcpPercentage           10
  TRtcpLimitScope           PORT_INSTANCE
}
IDSTRCondition              TRTcpCondition
{
  Protocol                  Tcp
  LocalPortRange            1:1023
}

#---------------------------------------------------------------------
# TR - IDSAction
#---------------------------------------------------------------------
IDSAction                   TRTcpLimit-action
{
   ActionType               TR LIMIT
   IDSReportSet             TRTcpLimitReportSet
   {
      TypeActions           Log
      TypeActions           Statistics
      StatType              Exception
   }
}
IDSAction                   TRTcpLog-action
{
   ActionType               TR NOLIMIT
   IDSReportSetRef          LogStatReportSet
}

#---------------------------------------------------------------------
# IDSReportSet
#---------------------------------------------------------------------
IDSReportSet                LogStatReportSet
{
   TypeActions              Log
   TypeActions              Statistics
}

#---------------------------------------------------------------------
```

```
# IPTimeCondition
#------------------------------------------------------------------------
IpTimeCondition            Time1
{
  TimeOfDayRange           1-22
  DayOfWeekMask            0111110
}
```

If you are using LDAP to define policy, see "IDS TCP traffic regulation policy example" on page 1406.

Restriction: LDAP policy cannot be used to define TCP traffic regulation policy that specifies IPv6 addresses.

The following traffic regulation UDP rule is defined:

- TR UDP: Defines UDP queue size for the low port range.
  - This rule provides statistics reports to determine normal traffic patterns for several applications while monitoring queue sizes.

The following example is an IDS configuration file:

```
###########################################################################
###########################################################################
# TR Policies
###########################################################################

#------------------------------------------------------------------------
# TR - IDSRule
#------------------------------------------------------------------------
IDSRule                    TRUdp-rule
{
  ConditionType            TR
  Priority                 2
  IDSTRConditionRef        TRUdpCondition
  IDSActionRef             TRUdpLogLimit-action
}

#------------------------------------------------------------------------
# TR - IDSTRCondition
#------------------------------------------------------------------------
IDSTRCondition             TRUdpCondition
{
  Protocol                 Udp
  LocalPortRange           1-1023
  TRUdpQueueSize           Long
}

#------------------------------------------------------------------------
# TR - IDSAction
#------------------------------------------------------------------------
IDSAction                  TRUdpLogLimit-action
{
   ActionType              TR LIMIT
   IDSReportSetRef         LogStatReportSet
}

#------------------------------------------------------------------------
# IDSReportSet
#------------------------------------------------------------------------
IDSReportSet               LogStatReportSet
{
   TypeActions             Log
   TypeActions             Statistics
}
```

If you are using LDAP to define policy, see "IDS UDP traffic regulation policy example" on page 1408.

Restriction: LDAP policy cannot be used to define UDP traffic regulation policy that specifies IPv6 addresses.

Verification

To verify that policies are correctly defined and functioning properly, consider the following points:

- Are the policies active?
- Is the expected traffic mapping to the correct policies?
- Are the IDS Policy functions working correctly?

The following subtopics provide more details about these considerations.

Are the correct policies active?

If your policies are defined in the LDAP server, check your LDAP server log or command output for errors encountered when your policies were loaded into LDAP. Some LDAP servers treat consecutive blank lines in an LDIF file as end of file; ensure that all of the policy objects in your LDIF files are acknowledged by LDAP.

Check your Policy Agent log file for errors while processing your policy.

Use the pasearch command to verify that the intended policies are active and have the expected attributes for the target stack.

Is the expected traffic mapping to the correct policies?

Use the Netstat IDS/-k SUMmary command to ensure that the intended policy has been mapped for each of the attack types, the Scan-Global type, the Scan-Event type for protocol ICMP, and the Scan-Event type for protocol ICMPv6. See z/OS Communications Server: IP System Administrator's Commands for more information on the Netstat command.

These IDS functions each select the single highest priority policy for their respective types at each policy change.

Use the Netstat IDS/-k PROTOcol TCP and Netstat IDS/-k PROTOcol UDP commands to ensure that the intended Scan-Event and TR policies have been mapped to the intended local sockets.

These IDS functions select the highest priority policy for the Protocol, local Port and local IP address when there is relevant activity against the socket.

For TCP this usually entails either a listen or the completion of an inbound connection handshake. For UDP this usually entails either a bind or an inbound datagram. Scan policies are also selected on some inbound error paths.

Are the IDS policy functions working correctly?

IDS policies that include IDS actions with statistics, log, or syslog set cause the stack to make log record information available to TRMD. If TRMD is running you may run the IDS report generator trmdstat against the appropriate log files to produce reports on the area of interest.

TRMD

TRMD runs as an APF-authorized program. The user ID associated with TRMD must be defined with a UID of 0, or must be permitted to become a superuser by having READ access to the BPX.SUPERUSER resource in the FACILITY class. See the EZARACF member of SEZAINST for sample RACF commands for TRMD.

Use the -p start option or the resolver configuration file to determine the stack that TRMD uses. Ensure that you specify the -p start option or that the RESOLVER_CONFIG environment variable is correctly set before starting TRMD. A separate instance of TRMD must be run for each TCP/IP stack.

The Log records written by TRMD contain two timestamps:

- A timestamp generated when the event was detected by the stack. This timestamp is generated by the stack and is always Coordinated Universal Time (UTC).
- A timestamp that is generated when the syslogd record ID is created. This timestamp is dependent on the setting of the TZ environment variable at the time that TRMD is started. If you want this timestamp

to be based on UTC, then ensure that the TZ environment variable is properly set (for example, export TZ=0) before starting TRMD.

You can set the TZ environment variable in the following ways:

- When starting TRMD from the z/OS shell:

  Export the TZ environment variable before starting TRMD; you should do this in /etc/profile or in .profile in the HOME directory. For example, if you are in the Eastern time zone in the United States:

  ```
  export TZ=EST5EDT
  ```

- When starting TRMD as a started task, use either of the following methods:
  - Specify TZ using the ENVAR parameter on the PARM statement in the started procedure. For example:

    ```
    // PARM='ENVAR("TZ=EST5EDT")/'
    ```

  - Export the TZ environment variable in a file specified with the STDENV DD statement. For example:

    ```
    // PARM='ENVAR("_CEE_ENVFILE=DD:STDENV")/'
    //STDENV DD PATH='/etc/trmd.env',PATHOPTS=(ORDONLY)
    ```

    Place the following statement in the /etc/trmd.env file:

    ```
    TZ=EST5EDT
    ```

    The use of the STDENV DD statement works well when you want to specify more than one environment variable; there is a JCL limit of 100 characters on the PARM parameter. Language Environment recommends a variable record format for the STDENV file.

You can also set the TZ environment variable for all applications in the CEEPRM*xx* PARMLIB member. You should define the TZ environment variable for all three LE option sets (CEEDOPT, CEECOPT, and CELQDOPT). For example:

```
CEECOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
CEEDOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
CELQDOPT(ALL31(ON), ENVAR('TZ=EST5EDT') )
```

For more information on specifying run-time options, see z/OS Language Environment Programming Guide. For details on setting the TZ environment variable, see z/OS UNIX System Services Command Reference.

If running multiple instances of TRMD, consider using the syslogd `-u` option when starting syslogd. The `-u` option causes the jobname of the application writing the log record to be included in the log record.

The TCP/IP stack must be running before TRMD can be started.

TRMD can be started from the z/OS shell or as a started task.

Running TRMD as a started task

A sample procedure is shipped in member EZATRMDP in SEZAINST. Follow the instructions in the sample member to define your environment.

The offset from Coordinated Universal Time (UTC) of the syslog time in the timestamp of TRMD messages is determined by the TZ environment variable. If the timestamp is required in UTC and has not been set by the TZ environment variable, specify the following definition in the TRMD procedure:

```
//  PARM=('POSIX(ON) ALL31(ON)',
//  'ENVAR("LIBPATH=/usr/lib"',
//  '"TZ=0")/-d 1')
```

To start TRMD as a started task, use the S TRMD command from the MVS console or SDSF. TRMD issues a fork, and in some cases the job name will be the original job name with a number appended. For

example, S TRMD might result in the TRMD started task running under the job name TRMD1, whereas S TRMDTASK would result in the TRMD started task running under the job name TRMDTASK. Use the D A,TRMD* command to verify the job name that TRMD is running under.

If running as a started task, issue P *jobname* to stop TRMD.

To automatically start TRMD when the TCP/IP stack is started, add TRMD to the AUTOLOG statement in the TCP/IP profile as follows:

```
AUTOLOG
    TRMD  JOBNAME TRMD
ENDAUTOLOG
```

Running TRMD from the z/OS UNIX shell

Ensure that you specify the -p start option or that the RESOLVER_CONFIG environment variable is correctly set before starting trmd.

The offset from Coordinated Universal Time (UTC) of the syslog time in the timestamp of TRMD messages is determined by the TZ environment variable. If the timestamp is to appear in Coordinated Universal Time (UTC), change the TZ specification in /etc/profile or export TZ="0" before starting TRMD.

After the appropriate environment is set up, issue the following command to start TRMD:

```
trmd
```

Stopping TRMD

To stop TRMD, issue the following kill command :

```
kill -s TERM pid
```

where pid is the TRMD process ID

To obtain the TRMD process ID, issue the following z/OS UNIX command:

```
ps  -A
```

Debug options can also be specified when starting TRMD. See z/OS Communications Server: IP Configuration Reference for more information.

trmdstat

The trmdstat utility program runs from the z/OS UNIX shell. The trmdstat program reads a log file, analyzes the log records generated by TRMD, and provides summary or detailed reports based on the options specified.

The following reports can be requested:

- IDS summary of logged events
- Reports of logged connection events
- Reports of logged intrusions defined in the ATTACK policy
- Reports of logged intrusions defined in the TCP policy
- Reports of logged intrusions defined in the UDP policy
- Reports of statistics events

See z/OS Communications Server: IP System Administrator's Commands for the trmdstat command and samples of the reports generated by the trmdstat program.

Defensive filtering

An external security information and event manager, by analyzing and correlating messages from multiple sources and systems in the network, can take action to block attacks by installing defensive filters in your TCP/IP stack. A defensive filter is a rule to discard packets, and is separate from IP security filters. Filter processing matches a defensive filter rule to data traffic based on any combination of IP source or destination address, protocol, source or destination port, or direction of flow. Filter processing checks defensive filters before IP security filters.

The z/OS UNIX **ipsec** command is used to add and manage defensive filters. Defensive filters are typically added as an automated action that results from the analysis of the external security information and event manager. However, you can also add a defensive filter by manually issuing the **ipsec** command. The Defense Manager Daemon (DMD) is an integral part of managing the defensive filters.

Figure 41 on page 185 shows an overview of defensive filtering and the DMD.

For more information about defensive filters and the DMD, see Chapter 19, “Defensive filtering,” on page 1107.

Chapter 17. IP security

This topic contains a description of IP filtering, IPSec-protected traffic, and preparing and configuring a z/OS system for IP security. Various business configurations are explained, including host-to-host, host-to-gateway, gateway-to-gateway, and gateway-to-host.

Terms and concepts for IP security

The following terms and concepts are used in this information:

3DES
Also known as triple DES, this encryption method uses three DES operations on a single data block with three different keys. Provides greater security than single DES.

Active
Used in three ways:

- Describes the filter policy that is in effect (default or Policy Agent).
- Describes the state of the rules or actions that are defined in Policy Agent. These rules or actions can be active or inactive due to a time condition.
- Describes the state of a manual tunnel installed in the TCP/IP stack. A manual tunnel can be active (available for use) or inactive (not available for use).

Active IPSec policy
The policy that is in effect, either the default filter policy or the IP security filter policy.

Advanced Encryption Standard (AES)
A symmetric block cipher that can encrypt (encipher) and decrypt (decipher) information. IP security for z/OS Communications Server supports AES with a 128-bit key length.

AES Cipher Block Chaining (CBC) (AES_CBC) mode
The AES algorithm using the CBC mode. IP security for z/OS Communications Server supports AES_CBC with a 128-bit or 256-bit key length.

AES Galois Counter Mode (GCM)
The AES algorithm using Galois Counter Mode and with a 16-byte integrity check value (ICV). Galois Counter Mode is a combined-mode algorithm that performs both encryption and authentication simultaneously. IP security for z/OS Communications Server supports AES_GCM with a 128-bit or 256-bit key length.

AES Galois Message Authentication Code (GMAC)
The AES algorithm using Galois Counter Mode to encode authentication data in either AH or ESP headers. AES_GMAC functions as a combined-mode algorithm; however, it provides authentication without encryption. IP security for z/OS Communications Server supports AES_GMAC with a 128-bit or 256-bit key length.

AES Extended Cipher Block Chaining (XCBC)
The AES algorithm using the XCBC mode to encode authentication data in either AH or ESP headers, with 128-bit keys and hash truncation to 96 bits.

Asymmetric encryption
Also known as public/private key encryption, this type of encryption is performed between two parties using pairs of encryption and decryption keys.

Authentication Header (AH)
An IP protocol (51) used with an IPSec Security Association to provide authentication of IP packets.

Autoactivation
The process by which a dynamic tunnel is activated when IP security policy is installed into the TCP/IP stack, either as the result of a user action, or the result of TCP/IP or the IKED initialization.

© Copyright IBM Corp. 2000, 2022

Certificate authority
A trusted third party that verifies information that is contained in an X.509 digital certificate.

Certificate revocation list (CRL)
A time-stamped list of revoked certificates that is signed by a certificate authority.

Child Security Association
The IKEv2 name for a phase 2 Security Association.

Command-line activation
The process of activating a tunnel through the use of the **ipsec** command. Both manual and dynamic tunnels can be activated from the z/OS UNIX command line.

CRLDistributionPoints
An optional X.509 certificate extension that identifies one or more locations where the CRL for a certificate is.

Data encryption standard (DES)
A block cipher with 64-bit blocks and a 56-bit key.

Default IP filter policy
Used until the IP security filter policy is installed by Policy Agent. The default IP filter policy includes both the filter rules you define in the TCP/IP profile and the implicit default filter rules that the stack generates. The implicit default filter rules deny all traffic that does not match any configured filter rule.

Dynamic tunnel
An IPSec tunnel whose security parameters are negotiated and whose encryption keys are generated dynamically using IKE.

Elliptic curve digital signature algorithm (ECDSA)
The algorithm that is used to authenticate a remote security endpoint using ECDSA with either SHA2-256 on the P-256 curve, SHA2-384 on the P-384 curve, or SHA2-521 on the P-521 curve.

Encapsulating Security Payload (ESP)
An IP protocol (50) used with an IPSec Security Association to provide authentication and encryption of IP packets.

Hashed message authentication code (HMAC)
A one-way hash function that combines the contents of a message and a secret key to produce a hash value; used for authentication.

HMAC_MD5
HMAC using the MD5 algorithm.

HMAC_SHA1
HMAC using a SHA1 algorithm that encodes authentication data in AH or ESP headers, using a 160-bit hash value and 96-bit integrity check value (ICV).

HMAC_SHA2
HMAC using a SHA2 algorithm that encodes authentication data in AH or ESP headers and that is qualified by the length of key and hash truncation. The algorithm can have 256-bit keys and hash truncation to 128 bits, 384-bit keys and hash truncation to 192 bits, or 512-bit keys and hash truncation to 256 bits.

IKE negotiation
A process by which two communicating IKE-enabled peers agree on a set of parameters that are used to protect traffic between them. This set of parameters is collectively known as a Security Association. One peer acts as the initiator of the negotiation, the other as the responder.

IKE Security Association
The IKEv2 name for a phase 1 Security Association.

IKE tunnel
A tunnel that protects IKE phase 2 messages.

Internet Key Exchange (IKE)
A protocol for the secure generation and management of encryption keys over an existing IP network. There are two versions, commonly referred to as IKE version 1.0 (IKEv1) and IKE version 2.0 (IKEv2).

Internet Security Association and Key Management Protocol (ISAKMP)
Defines IKEv1 procedures and packet formats to establish, negotiate, modify, and delete Security Associations.

IP filter rule
A configured rule that defines the action applied to an IP traffic pattern that is encompassed by the rule. The possible actions include permit, deny, and permit with IPSec protection.

IP filter table
An ordered list of IP filter rules. When IP filtering is active on a host, the table is consulted for each IP packet that is sent or received. The action of the matching IP filter rule is enforced by the TCP/IP stack.

IPSec
A suite of protocols and standards defined by the Internet Engineering Task Force (IETF) for secure communication over an existing IP network.

IPSec tunnel
A tunnel that protects IP traffic between two endpoints using one or both of the IPSec protocols. Manual and dynamic tunnels are both instances of an IPSec tunnel.

IP security filter policy
The policy that is installed by the Policy Agent. It includes the filter rules you define in the Policy Agent configuration files and an implicit deny all rule that is generated by Policy Agent.

IP traffic pattern
The set of IP traffic attributes that can be used as input to an IP filter table query. Typically, this includes IP source address, IP destination address, source port, destination port, protocol, and direction (inbound or outbound).

Manual tunnel
An IPSec tunnel whose security parameters and encryption keys are statically configured and must be manually managed by a security administrator.

Message authentication code (MAC)
A tag derived from the contents of a message and a secret key. The tag can be used to authenticate the integrity of a message as well as the source of the message.

Message digest algorithm 5 (MD5)
A MAC algorithm that produces a 128-bit hash value.

NAT traversal (NATT)
Traversal of IPSec traffic through a NAT device.

Network address port translation (NAPT)
A technique where multiple internal IP addresses are translated into a single public IP address. As part of this translation process, the TCP and UDP ports in the packets are translated. NAPT is sometimes referred to as port address translation (PAT) or IP masquerade.

Network address translation (NAT)
Network address translation is a broad term that encompasses both a one-to-one address translation function, translating a single internal IP address to a single public IP address, and the NAPT function.

Network security services (NSS)
Services performed in support of security enforcement or management.

NSS client
Requests network security services from an NSS server. The z/OS IKE daemon can act as an NSS client for a TCP/IP stack.

NSS server
Provides network security services for one or more NSS clients.

On-demand
The process by which a dynamic tunnel is activated by outbound traffic flow without user intervention.

Phase 1
The first stage of an IKE negotiation, in which an ISAKMP Security Association is established between two IKEv1-enabled peers, or in which an IKE Security Association is negotiated between two IKEv2-

enabled peers. A phase 1 Security Association refers to IKEv1 ISAKMP SAs, as well as to IKEv2 IKE SAs.

Phase 2
The second stage of an IKE negotiation, in which an IPSec Security Association is established between two IKEv1-enabled peers, or in which a child Security Association is negotiated between two IKEv2-enabled peers. A phase 2 Security Association refers to IKEv1 IPSec SAs, as well as to IKEv2 child SAs.

Rivest Shamir Adleman (RSA)
An asymmetric key encryption method, in which the key that is used to encrypt data is different than the key that is used to decrypt the data. RSA can be used for encryption, or to authenticate a digital signature.

Secure hash algorithm 1 (SHA1)
A MAC algorithm similar to MD5, but more secure. This algorithm produces a 160-bit hash value.

Secure hash algorithm 2 (SHA2)
A MAC algorithm similar to SHA1, but more secure. This algorithm produces a 256-bit, 384-bit or 512-bit hash value.

Security Association (SA)
An agreement between two IPSec-enabled hosts that describes the type of data to protect and the methods that are used to protect the data. IKE creates a phase 1 Security Association to protect IKE messages (also known as the ISAKMP Security Association or the IKE Security Association), and a phase 2 Security Association to protect data traffic (also known as the IPSec Security Association or the child Security Association).

Symmetric encryption
Encryption that is performed between two parties sharing the same encryption key. Also known as secret key encryption.

Transport mode encapsulation
A process used to construct IPSec packets by inserting one or more additional IPSec headers between the IP header to be protected and the IP payload of the packet to be protected.

Tunnel
A secure logical connection or channel that is defined by a collection of Security Associations that define the security parameters protecting traffic between two endpoints.

Tunnel activation
The process by which a tunnel becomes active or usable. For dynamic tunnels, this process involves initiating an IKE negotiation.

Tunnel mode encapsulation
A process used to construct IPSec packets by creating a new IP header with an IP payload consisting of the entire IP packet being protected, and then inserting one or more additional IPSec headers between the new IP header and its IP payload (that is, the original IP packet).

UDP encapsulation
A process used to construct IPSec packets by first applying tunnel mode encapsulation or transport mode encapsulation to an IP packet to be protected by the ESP protocol, and then inserting a UDP header between the IP header and the ESP header.

Virtual private network (VPN)
A logical network of connected network nodes that communicate through secure channels (tunnels), typically by using the IPSec protocols (AH and ESP).

X.500 distinguished name
A collection of X.509 values, such as common name, host name, organization, organizational unit, and so on, that is stored in an X.509 digital certificate. An X.500 distinguished name is used as a globally unique identifier for the owner.

X.509 digital certificate
A set of information in the X.509 standard containing various attributes about an entity, including identity information and a public key that is used for encrypted communications with that entity.

Terminology conventions for IP security

The following terminology conventions are used throughout this information when referring to z/OS IP security:

IP security
The z/OS Communications Server function.

IPSec
The protocol suite.

ipsec
The action associated with an IP filter action, or the z/OS UNIX System Services command.

IPSEC
The statement in the TCP/IP profile.

IPSECURITY
The parameter on the IPCONFIG statement in the TCP/IP profile.

NAT
The general network address translation function. NAT encompasses both one-to-one address translation and network address port translation.

NAPT
Network address port translation. This term is used when information is specific only to this form of NAT.

Commands used to administer IP security

The following commands are used to administer IP security. For more information on these commands, see z/OS Communications Server: IP System Administrator's Commands.

certbundle
Use the z/OS UNIX System Services **certbundle** command to create a certificate bundle file that contains certificate and CRL information.

ipsec
Use the z/OS UNIX System Services **ipsec** command to display information about active filters and Security Associations, and to control aspects of Security Association negotiation. The **ipsec** command is used to:

- Display filters that are active in the stack
- Revert to default IP filter policy, as defined in the TCP/IP profile
- Reload IP security policy, as defined in the Policy Agent configuration files
- Activate Security Association negotiations
- Display existing phase 1 Security Associations
- Display existing phase 2 Security Associations
- Display remote port mappings used with NAT traversal configurations
- Display network security configuration information for the active stacks on the local system
- Display information for each NSS IPSec client that is currently connected to the NSS server
- Refresh existing phase 1 Security Associations
- Refresh existing phase 2 Security Associations
- Deactivate existing phase 1 Security Associations
- Deactivate existing phase 2 Security Associations
- Test for a filter rule match for a given set of IP traffic characteristics

Authority to use the **ipsec** command is controlled through RACF. There are two distinct types of SERVAUTH profiles that define access to the **ipsec** command, one for display capabilities and one for control capabilities.

Tip: Many of the tasks, examples, and references in this information assume that you are using the z/OS Security Server (RACF). References to RACF apply to any other SAF-compliant security products that contain the required support. If you are using another security product, read the documentation for that product for instructions on task performance.

For the steps to configure access control to the **ipsec** command, see Appendix E, “Steps for preparing to run IP security,” on page 1361.

For detailed syntax and usage, and how to control access of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

pasearch

Use the **pasearch** command to display Policy Agent information that is defined in the Policy Agent configuration files, including IP security and other types of policies. The options that are related to IP security include the ability to view IP security policy rules and actions, both active and inactive, for any TCP/IP stack for which policies have been defined and that is IPSECURITY-enabled.

If the user of the **pasearch** command is not a superuser, authority is controlled through RACF.

For detailed syntax and usage of the **pasearch** command, see z/OS Communications Server: IP System Administrator's Commands.

MODIFY

Use the MODIFY console command to have:

- The IKE daemon reread the IKED configuration file
- Policy Agent reread the Policy Agent configuration files

For detailed syntax and usage of the MODIFY command, see z/OS Communications Server: IP System Administrator's Commands.

Netstat

Use the Netstat command to display the following information:

- IPSECURITY enablement for a particular stack (Netstat CONFIG/-f)
- SecurityClass (SECCLASS) for a specific interface (Netstat DEVLINKS/-d)

For detailed syntax and usage of the Netstat command, see z/OS Communications Server: IP System Administrator's Commands.

Overview of using IP security

z/OS Communications Server provides the ability to control and monitor network traffic on one or more TCP/IP stacks on a z/OS system. IP security for z/OS Communications Server supports IP filtering, IPSec, and Internet Key Exchange (IKE). IP security for z/OS Communications Server supports two versions of the IKE protocol: IKEv1 and IKEv2. See “Dynamic key management - IKE and IPSec negotiations” on page 926 for more information.

IP security policy can be used for the following protection:

- Protect a secure host on an internal network from unwanted network traffic
- Provide protection for traffic between partner companies over connected networks
- Allow secure sending of data over the Internet by providing IPSec virtual private network (VPN) support

These features are implemented in the IP layer on a per packet basis, and thus are available to any network application without requiring any special modifications. Applications can also implement their own additional security features as necessary, on top of the underlying IP security.

IP security policy is enabled, enforced, managed, and monitored through a coordinated effort of several z/OS Communications Server components:

- Policy Agent

The Policy Agent is used to configure IP security on a z/OS system. It reads the configuration files that contain the IP security policy configuration statements, checks them for errors, and installs them into the IKE daemon and the TCP/IP stack.

- Internet Key Exchange daemon (IKED)

  The Internet Key Exchange daemon is responsible for retrieving IP security policy from Policy Agent, and dynamically managing keys that are associated with dynamic IPSec VPNs. This daemon also provides network management capabilities for IP security aspects of local TCP/IP stacks.

- Network Security Services daemon (NSSD)

  The Network Security Services daemon provides the NSS server functionality. It provides the NSS IPSec certificate service to perform digital signature creation and verification operations on behalf of NSS IPSec clients. The NSS IPSec certificate service is used by the IKED during phase 1 negotiations when digital signature authentication is required.

- TCP/IP stack

  The stack maintains a list of currently active IP filters and IPSec Security Associations, actively filters network traffic, controls encryption and decryption of network data, and maintains counters that are associated with an IPSec Security Association lifetime.

- Traffic Regulation Manager daemon (TRMD)

  The Traffic Regulation Manager daemon is responsible for logging IP security events that are detected by the stack, including IP filter events, updates to IP security policy, and the creation, deletion, and refresh of IPSec Security Associations.

- System logging daemon (syslogd)

  The system logging daemon manages the logging of messages and events for all of the other components, including where the log messages are written.

These components provide a combination of technologies that form the basis of IP security:

- IP filtering
- IP filter logging
- Data encryption and authentication

FIPS 140 mode and IP security

The Internet Key Exchange daemon (IKED), the network security services daemon (NSSD), and the TCP/IP stack components perform a wide variety of cryptographic operations for IP Security. The IKED and the NSSD manage cryptographic keys and digital signatures and perform hashes for authentication. The IKED and the TCP/IP stacks perform encryption and decryption of messages that flow over the IPSec tunnels. Architectural enhancements to the IKE protocols periodically introduce new cryptographic algorithms for performing hashes and encryption.

Federal Information Processing Standards (FIPS) document 140 (FIPS 140) provides a higher degree of assurance of the integrity of cryptographic operations by placing restrictions on the cryptographic components and the operations performed by these components. Weaker algorithms are forbidden and all the operations must be performed by cryptographic modules that are contained within a well defined cryptographic boundary.

You can configure the IKED, the NSSD, and the TCP/IP stack components to operate in FIPS 140 mode. When you do, you restrict the cryptographic algorithms they support, and you modify their interactions with each other and with the other hardware and software components of the z/OS system related to cryptography, such as Integrated Cryptographic Services Facility (ICSF) and System SSL.

In FIPS 140 mode, the IKED, the NSSD, and the TCP/IP stacks enforce the following restrictions on the cryptographic algorithms that can be used for IP security:

- You cannot use the DES encryption algorithm.

- You cannot use the HMAC-MD5, HMAC-MD5-96, AES128-XCBC, and AES128-XCBC-96 algorithms for authentication or pseudo-random function.
- You cannot use Diffie-Hellman groups 1, 2, and 5.
- You cannot use certificates for RSA signature authentication that have key lengths less than 1024 bits.
- You cannot use pre-shared keys whose length is less than half the key size of the chosen HowToAuthMsgs (IKEv1) or PseudoRandomFunction (IKEv2) algorithm.

When the FIPS 140 mode is configured for a TCP/IP stack, the Policy Agent enforces some of the FIPS 140-related restrictions when it parses the IP security policy files. Other restrictions are enforced when dynamic tunnels are being activated, after the FIPS 140 mode of all of the relevant software components (the IKED and the NSSD) is known.

You configure FIPS 140 mode independently for each of the IKED, the NSSD, and the TCP/IP stack components. ICSF and System SSL must also be configured to support FIPS 140. If you use FIPS 140 mode in some components and not others, the resulting system might not operate in FIPS 140 mode. The components that are configured to use FIPS 140 mode can use cryptographic services only from components that are also operating in FIPS 140 mode, so the FIPS 140 mode mismatch can cause functional problems.

- If a TCP/IP stack is configured to use FIPS 140 mode, but the IKED is not, the IKED does not perform any dynamic VPN tunnel activations for the stack. The IKED performs many cryptographic operations on behalf of the stack during tunnel activation.

  Rule: Whenever the stack is configured for FIPS 140 mode, also configure FIPS 140 mode for the IKED.
- If the IKED and the TCP/IP stacks it supports are configured to use FIPS 140 mode, but the NSSD is not, the IKED cannot use the NSS IPSec certificate service for its stacks. Because the IKED must use the NSS IPSec certificate service to create and verify digital signatures for IKEv2 tunnels, this FIPS 140 mode mismatch prevents activation of any IKEv2 tunnels that use digital signature authentication.

  Rule: Whenever the IKED is configured for FIPS 140 mode, also configure FIPS 140 mode for the NSSD.
- If you have a sysplex that is configured for Sysplex-Wide Security Associations (SWSA), and the distributor stack is not configured to use FIPS 140 mode, then it will not be able to successfully distribute tunnels to target stacks that are configured in FIPS 140 mode.

  Rule: Whenever you have target stacks that are configured in FIPS 140 mode, first configure FIPS 140 mode for the distributing stack.

When possible, you should enable FIPS 140 mode for the IKED, the NSSD, and the TCP/IP stacks all at once. If you must implement FIPS 140 support in stages, enable FIPS 140 mode in the components in the following order:

1. Configure FIPS 140 mode in the NSSD. If the NSSD is configured in FIPS 140 mode and the IKED and the TCP/IP stacks are not, the IKED still uses the NSS IPSec certificate service provided by the NSSD. Note that the NSSD creates and verifies signatures only for certificates that conform to FIPS 140 restrictions, even if the IPSec client is not operating in FIPS 140 mode.
2. Configure FIPS 140 mode in the IKED. When the NSSD and the IKED are both in FIPS 140 mode, but the stacks are not, dynamic VPN tunnels can be activated and data can flow, as long as those tunnels follow the FIPS 140 cryptographic algorithm restrictions.
3. If you are using SWSA in a sysplex, configure FIPS 140 mode in the distributor stack of the sysplex.
4. Configure FIPS 140 mode in all other TCP/IP stacks.

Enabling FIPS 140 mode on a system can affect performance. For example, you might have to change from using a weak encryption algorithm to using one that requires more processing to perform. Even if no algorithm changes are necessary, the IKED, the NSSD, and the TCP/IP stacks perform their cryptography operations in a different way when FIPS 140 mode is enabled than when it is not enabled, because FIPS 140 imposes additional self-verification requirements and access restrictions, and because hardware accelerated implementations of some cryptographic operations might not be available in FIPS 140 mode.

Steps for configuring IP security to support FIPS 140 mode

Configure IP security to support FIPS 140 mode on each system and stack that needs to use FIPS 140 mode. If you are using Sysplex-Wide Security Associations (SWSA), perform these steps first on your distributor and backup stacks, and then on each of your target stacks.

Procedure

Perform the following steps to configure IP security to support FIPS 140 mode:

1. Ensure that Integrated Cryptographic Services Facility (ICSF) is started and configured to support FIPS 140.

   Requirement: ICSF must be active before starting the IKE daemon or NSS server configured in FIPS 140 mode. For information about enabling ICSF to support FIPS 140-2, see the topic about operating in compliance with FIPS 140-2 in z/OS Cryptographic Services ICSF Writing PKCS #11 Applications.

   Tip: You do not need to create TKDS data sets in order for IP security to use ICSF.

   For more information about enabling FIPS 140 mode for ICSF, see z/OS Cryptographic Services ICSF Administrator's Guide.
2. Ensure that one of the following conditions are true:

   - The SAF class CRYPTOZ is not active.
   - No SAF profile exists for the FIPSEXEMPT.SYSTOK-SESSION-ONLY resource in the CRYPTOZ class.
   - The IKED, the NSSD, and the TCP/IP stacks that are configured in FIPS 140 mode have no access (NONE) to the SAF resource FIPSEXEMPT.SYSTOK-SESSION-ONLY in the CRYPTOZ class.

     Tip: A single z/OS system can support multiple TCP/IP stacks, and you can configure some TCP/IP stacks with FIPS 140 support and others without FIPS 140 support. The stacks that are configured in FIPS 140 mode must not have access to the SAF resource FIPSEXEMPT.SYSTOK-SESSION-ONLY in the CRYPTOZ class.
3. Ensure that System SSL FIPS 140 support is available and configured. For more information, see the information about System SSL and FIPS 140-2 in z/OS Cryptographic Services System SSL Programming.
4. If you are using network security services (NSS), configure NSS to support FIPS 140.

   You can configure FIPS 140 by specifying `Yes` as the FIPS140 value in the NSS configuration file (for example, nssd.conf). In the IBM Configuration Assistant for z/OS Communications Server, configure the FIPS 140 option in the Advanced Server Settings for NSS in the NSS perspective.

   After you have configured FIPS 140, restart the NSS daemon if it was active.

   Tip: If TCP/IP is enabled for FIPS 140 but the NSSD is not, then the NSSD cannot provide NSS certificate services to the TCP/IP stack.
5. Configure IKE to support FIPS 140.

   You can configure FIPS 140 by specifying `Yes` as the FIPS140 value in the IKED configuration file (for example, iked.conf). In the IBM Configuration Assistant for z/OS Communications Server, configure the FIPS 140 option in the Advanced IKE Daemon Settings in the IPSec perspective.

   After you have configured FIPS 140, restart the IKE daemon if it was active.

   Tip: If TCP/IP is enabled for FIPS 140 but the IKED is not, then the IKED will not negotiate dynamic VPN tunnels for that TCP/IP stack.
6. Configure the TCP/IP stack to support FIPS 140.

   You can configure FIPS 140 by specifying `FIPS140 Yes` on the IpFilterPolicy statement in the IPSec policy file for the stack. In the IBM Configuration Assistant for z/OS Communications Server, configure the FIPS 140 option in the Advanced Stack Settings in the IPSec perspective.

   After you have configured FIPS 140, restart the stack if it was active.

Configuring IP security

You configure z/OS IP security using an extensive set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the set of IP security requirements for each TCP/IP stack. IBM provides two alternatives for creating the Policy Agent configuration files:

- "Configuring IP security using the IBM Configuration Assistant for z/OS Communications Server" on page 892
- "Configuring IP security using manual configuration" on page 892

Configuring IP security using the IBM Configuration Assistant for z/OS Communications Server

The IBM Configuration Assistant for z/OS Communications Server, an optional GUI-based tool, provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the Policy Agent and IKE daemon configuration files.

The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

Through a series of wizards and online help panels, you can use the IBM Configuration Assistant for z/OS Communications Server to create IP security configuration files for any number of z/OS images with any number of TCP/IP stacks per image. Using the IBM Configuration Assistant for z/OS Communications Server, there are four types of reusable objects:

- Traffic descriptors that define the IP traffic type, such as TCP or UDP
- Security levels that define the different ways to protect data, such as the encryption level
- Requirement maps that map traffic descriptors to security levels

  A single requirement map should contain a complete set of security requirements that will govern the level of security for multiple IP traffic types.
- Address groups that define a set of addresses to be used in an IP filter rule

For each TCP/IP stack, you create a set of connectivity rules that indicate the data endpoints and indicate which requirement map will govern security between the data endpoints.

The IBM Configuration Assistant for z/OS Communications Server comes with a number of IBM-supplied traffic descriptors, security levels, and requirement maps that are easily applied to an existing network topology, or the IBM-supplied definitions can be used as the basis for your own set of reusable objects.

The IBM Configuration Assistant for z/OS Communications Server can dramatically reduce the amount of time that is required to create IP security policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS security, use of the GUI is encouraged to ensure that you have a consistent and easily manageable interface for implementing IP security.

This information primarily describes option 2, manual configuration. However, if you are using the IBM Configuration Assistant for z/OS Communications Server, reading this information will help you understand security concepts and the relationship between Policy Agent and IP security function.

Configuring IP security using manual configuration

You can manually create the IP security policy configuration files by coding all of the required statements in a z/OS UNIX file or MVS data set. There are a large number of configuration options provided by IP security policy statements that permit advanced users to carefully fine-tune IP security policy on a per-stack basis. This information describes the procedure for creating an IP security policy by manually creating and editing the configuration files. There are examples that step through the process of creating a configuration file that includes zones corresponding to various security models. For details about the

configuration statements and parameters, see the Policy Agent and policy applications topic in z/OS Communications Server: IP Configuration Reference.

Specifying the IP security configuration file based on Policy Agent role

The Policy Agent can act as a policy server, a policy client, or neither. For more information on these different roles, see "Policy types and infrastructure overview" on page 787. Regardless of which option is used to configure IP security policies, the resulting configuration files need to be specified using different statements, depending on the role of the Policy Agent.

- If you are using the Policy Agent as a policy client that retrieves IP security policies from the policy server, specify the configuration files using the DynamicConfigPolicyLoad statement on the policy server.
- If you are using the Policy Agent as a policy client, but the policy client does not retrieve IP security policies from the policy server, specify the configuration files using the IPSecConfig statement on the policy client.
- If you are not using a policy client/policy server environment, specify the configuration files using the IPSecConfig statement on the single Policy Agent.

When this information refers to configuration files, keep in mind where the files should exist, based on the role of the Policy Agent.

IP filtering

IP filtering controls the flow of network traffic. An administrator can deny or allow any given network packet into or out of a z/OS system with an IP security policy.

Filter rules and actions

The IP security policy enables a z/OS system to classify any IP packet that comes across a network interface and take specific action according to a predefined set of rules. The set of properties that identify a packet, together with the action to be performed on it, is known as an IP filter rule. The rule can be used to filter out unwanted packets from the network stream, while allowing others. The collection of all filter rules comprise the IP filter table. The IP filter table contains all of the IP filter rules in the order in which they were configured. IP filter rules are configured using the IpFilterRule statement in an IP security policy configuration file. For more details about the IpFilterRule statement, see z/OS Communications Server: IP Configuration Reference.

For example, a simple filter table might have the following set of rules:

1. Allow Telnet traffic from IP address A.
2. Allow FTP traffic from IP address B and IP address C.
3. Allow any traffic from subnet D, but ensure that it is encrypted.
4. Allow IPsec-protected traffic from any location if the remote IKE identity is a corporate email address.
5. Allow outbound connections to anywhere.
6. Deny anything that does not match the previous rules.

This set of rules would be considered to be a filter table consisting of six rules.

The filter table that is configured for a particular installation reflects the security needs for that site. The rules can be restrictive or permissive, as the security policy allows. Normally the rules would deny anything not explicitly permitted, a configuration known as a default-deny policy, in which rules are added as necessary to allow only crucial network traffic. In a default-deny environment, the absence of any IP filter rules essentially isolates the system from the network. The alternative, a default-allow policy, allows all network traffic in the absence of any configured rules. Specific rules can be added as needed to deny unwanted or potentially malicious traffic. A default-deny policy is considered to be much more secure.

Rule: A z/OS Communications Server TCP/IP stack that is configured for IP security follows a default-deny policy by default, in the absence of any configured filter rules.

IP filter tables can grow very complex, and in many implementations are difficult to maintain. However, the configuration mechanism that is provided by IP security enables you to attach meaningful descriptors to rules, hosts, and other configured items, which makes keeping track of complex filter tables an easier task.

To protect data between hosts, hosts must agree on what type of traffic to protect, and how to protect that traffic. These IP traffic pattern definitions are stored in the locally configured security policy and installed in the IP filter table, which is consulted for each IP packet that enters or leaves the system. When a packet matches one of the rules in the IP filter table, the policy determines what action is taken for that packet. IP filter actions are configured using the IpGenericFilterAction statement in an IP security policy configuration file. For more details about the IpGenericFilterAction statement, see z/OS Communications Server: IP Configuration Reference.

On a z/OS stack that has IPCONFIG IPSECURITY configured (and perhaps also has IPCONFIG6 IPSECURITY configured) and an active IP security policy, there are three possible actions:

- Deny the packet.
- Permit the packet.
- Permit the packet with IPSec protection.

If the action that is associated with the filter rule is an ipsec action, the packet is subject to the application of IPSec authentication and encryption before it is received or sent. Any packet that matches a filter rule with an ipsec action is processed using the IPSec protocols, either Authentication Header (AH), Encapsulating Security Payload (ESP), or both, depending on the locally configured policy. z/OS IP security requires that data authentication be done if the filter rule specifies an ipsec action.

Filtering criteria in an IP packet

IP packets match IP filters based on a number of selection criteria. There are five primary pieces of information that are gathered from the IP packet, commonly referred to as a 5-tuple:

- Source address

  An IP packet can be filtered based on the source address located in the IP header of the packet.
- Destination address

  An IP packet can be filtered based on the destination address located in the IP header of the packet. For IPv6 packets that contain a type 0 or type 2 routing header, the stack performs IP filtering using the final destination address of the packet based on the routing header contents, not on the destination address in the IP header.
- Protocol

  An IP packet can be filtered based on the protocol in the IP header of the packet.
- Source port

  If the protocol in an IP packet is TCP or UDP, the packet can be filtered based on the source port in the TCP/UDP header of the data portion of the packet.
- Destination port

  If the protocol in an IP packet is TCP or UDP, the packet can be filtered based on the destination port in the TCP/UDP header of the data portion of the packet.

Additional filtering criteria based on protocol

The following criteria are some additional filtering criteria that are based on protocol:

- ICMP type and code

  If the protocol in an IP packet is ICMP or ICMPv6, the packet can be filtered based on the ICMP type and code located in the ICMP header of the data portion of the packet.

Guideline: Typically, if an ICMP error is generated for a packet that arrived over a Security Association, and there is no matching rule for the ICMP error, z/OS Communications Server attempts to send the ICMP packet over the same Security Association that carried the original packet. If you want to disable this processing, you must code a rule that covers ICMP errors.

- OSPF type

  If the protocol in an IP packet is OSPF, the packet can be filtered based on the OSPF type located in the OSPF header of the data portion of the packet.

- Mobility header type

  If the protocol in an IP packet is IPv6 mobility header, the packet can be filtered based on the mobility header type located in the mobility header of the data portion of the packet.

Additional filtering criteria based on network attributes

Some filtering criteria are inferred from the external characteristics of the IP traffic, rather than being found in the actual IP packet. The following additional attributes are used to distinguish IP packets:

- Direction

  The direction of an IP packet is either inbound or outbound.

- Routing

  The routing attribute of an IP packet is either local or routed. IP packets are considered local if either of the following conditions is true:

  – The packet is inbound and the destination address in the IP packet exists on the IPSECURITY stack.

  – The packet is outbound and the source address in the IP packet exists on the IPSECURITY stack.

  Otherwise the IP packet is considered routed.

- Security class

  Each non-virtual interface on a z/OS system is assigned a security class. The security class of an IPv4 interface is determined by the SECCLASS parameter that is coded on the LINK statement or the DYNAMICXCF parameter of the IPCONFIG statement in the TCP/IP profile. The security class of an IPv6 interface is determined by the SECCLASS parameter that is coded on the INTERFACE statement or the DYNAMICXCF parameter of the IPCONFIG6 statement in the TCP/IP profile. The value of SECCLASS is a number in the range 1-255. If SECCLASS is not specified for an interface, the interface has a default security class of 255.

  Each IP packet entering or leaving the system inherits the security class of the interface that it traverses:

  – For inbound traffic, this is the interface on which the packet arrived.

  – For outbound traffic, this is the interface over which the packet is sent.

  The value for SECCLASS has no inherent meaning. A SECCLASS of 255 is no more or less secure than a SECCLASS of 1. You can optionally assign a SECCLASS value either to uniquely identify an interface, or to group interfaces with similar security requirements, based on site policy. Consequently, this attribute can be used as an additional criterion for IP filtering. Security class can be used to define broad filter rules that encompass all of the IP traffic that uses a group of interfaces without explicit knowledge of network addressing, or to ensure that an IP packet arrived on a valid interface.

  For example, as shown in Figure 110 on page 896, consider a z/OS system with three physical interfaces that are assigned the following security classes:

```
I1: SECCLASS 10
I2: SECCLASS 20
I3: SECCLASS 20
```

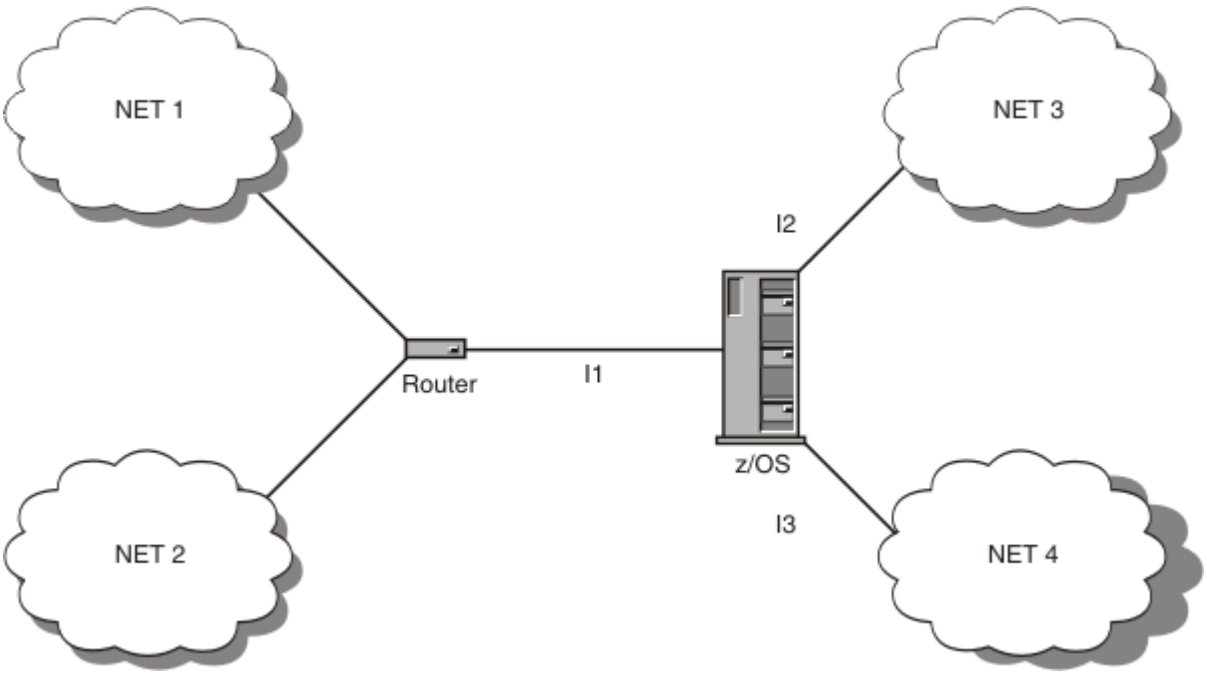


Figure 110. Using SECCLASS to identify interfaces

All IP packets from NET1 and NET2 have the same security class (10), and all IP packets from NET3 and NET4 have the same security class (20). Even though each interface on the z/OS system can reach multiple networks, you can configure a single IP filter rule that matches all of the IP traffic from NET3 and NET4 without explicitly identifying any attributes of the IP packets, other than security class. For instance, if NET3 and NET4 are trusted internal networks in which you want to allow all network traffic, you can configure one IP filter rule permitting all traffic with a security class of 20, without regard to IP address.

An IP filter using security class as a criterion is not limited to scenarios in which the IP addresses are ignored. To counter IP address spoofing, IP filter rules can take into account both security class and IP address. IP address spoofing involves the creation of an IP packet whose source address has been modified to reflect an address other than that of the originator. Because routers ignore the source address when making routing decisions, the modified IP packet can still reach its destination. An IP packet whose source IP address has been spoofed is usually not legitimate and is often used in a malicious manner. IP filtering can counter an attempt to spoof a source IP address by ensuring that an IP packet arrived on a valid network interface. For instance, any inbound packet that has a source address of an internal network node, but entered the system though an external interface, is probably spoofed and should be denied.

Rule: The SecurityClass parameter cannot be used on an IpFilterRule statement with a dynamic ipsec action. The IKE protocols do not provide for the negotiation of SecurityClass, and therefore it cannot be used as a selection criterion for a Security Association. The SecurityClass parameter can be configured only for an IpFilterRule statement that has an action of permit, deny, or manual ipsec. Coding a SecurityClass parameter with a value other than 0 on a dynamic ipsec rule is considered a configuration error.

Guidelines:

- If there are multiple interfaces by which a packet can reach its destination, the SECCLASS values for the interfaces and the filter rules should be consistent. If a packet could reach its intended destination by traversing one of a number of interfaces, the SECCLASS values and the related filter rules should account for the possibility of alternate routing.
- If the same IPv6 link-local address exists in multiple LANs in your network and you want to define filter rules to distinguish these IP addresses, configure different SECCLASS values for the interfaces onto these LANs to make the filter rules unique. However, if you are using dynamic IPSec for link-

local addresses, you cannot use SECCLASS values to make the filter rules unique, so you must ensure that there are no overlapping or duplicate link-local addresses.

For more details about the SecurityClass parameter on the IpService statement, the SECCLASS parameter on the LINK and INTERFACE statements, and the SECCLASS keyword on the DYNAMICXCF parameter of the IPCONFIG and IPCONFIG6 statements, see z/OS Communications Server: IP Configuration Reference.

- Remote IKE identity

  For traffic that is protected by IPSec, the remote IKE identity can be used as a filtering criterion. This is particularly useful in cases where a client is roaming or mobile and does not have a fixed IP address, but has a known authenticated IKE identity.

  Filter rules with a remote IKE identity often have wildcard IP addresses that are not very specific, because mobile users can connect from a variety of locations.

  Guidelines:

  – Consider placing filter rules with a remote IKE identity near the bottom of your IP filter policy, so that other traffic has optimal filter searching. Because the wildcard IP addresses are not very specific for these rules, much IP traffic needs to be compared against these rules to test whether the traffic matches the rule.

  – Consider whether the mobile client's traffic might intersect with other filter rules; for example, when a mobile user occasionally brings their laptop into the office. If you choose to permit the user to continue to establish their IPSec connection while in the office, then you should order the remote identity filter rules in your IP filter policy so that they precede the office network filter rules.

IP traffic patterns

The information that is gathered from an IP packet can be used to identify a specific type of network traffic, based on common Internet protocols. For example:

- Web traffic to a server consists of TCP packets inbound to and outbound from port 80.
- TN3270E Telnet server traffic consists of TCP packets inbound to and outbound from port 23.
- IKE traffic consists of UDP packets inbound from and outbound to port 500 or 4500.
- Ping (IPv4) consists of ICMP packets outbound with type 8, code 0 (echo request), and inbound with type 0, code 0, (echo reply).
- Ping (IPv6) consists of ICMPv6 packets outbound with type 128, code 0 (echo request), and inbound with type 129, code 0, (echo reply).

The IpService statement groups these attributes into definitions of common traffic patterns that can subsequently be incorporated in an IP filter rule. The sample configuration file /usr/lpp/tcpip/samples/pagent_CommonIPSec.conf provides definitions for many of the common traffic patterns that are seen on a typical z/OS server.

For more details about the IpService statement, see z/OS Communications Server: IP Configuration Reference.

Routed traffic and fragmented packets

Fragmented packets are problematic for IP security implementations because transport layer headers (UDP, TCP, and so on) are not present in every fragment. For some packet fragments, the transport layer selector values (for example, UDP ports, ICMP type and code) are indeterminate, which complicates IP packet filtering. In the case of IPv6, a fragment's IP protocol value might also be indeterminate.

Tip: z/OS Communications Server filters fragmented packets only for routed traffic. For all local traffic, z/OS Communications Server performs filtering on the fully assembled packet.

Because some packet fragments do not contain the transport layer header and cannot be filtered based on transport layer selector values, z/OS Communications Server enforces the following restriction for routed traffic.

Restriction: All routed traffic between two endpoints must be filtered in the same manner without regard to TCP or UDP port, ICMP or ICMPv6 type and code, or OSPF and MIPv6 type.

Guideline: To perform more granular filtering for this traffic, enforce a port-specific filtering policy, or a type-specific and code-specific filtering policy, at the final destination for this IP traffic.

z/OS Communications Server supports the specification of Opaque for matching indeterminate IPv6 protocol values in packet fragments. Packets that have an indeterminate value match filter rules only when Opaque or Any is specified.

z/OS Communications Server also supports explicit matching on fragmented packets. The IpService policy object's fragment specification can be used to match any fragmented packet, even if the packet's selector values are known. Explicit matching can be used to deny all fragmented packets in situations in which they are not part of normal traffic patterns and are considered to be a security risk.

Conditionally controlling IP filters

When IP filters from IP security policy files are installed into the stack, they are always active by default. Depending on the business need, this might not always be desirable. For instance, you might want to restrict certain types of IP traffic to a specific time, day, or week. IP security provides this flexibility by allowing you to specify a time condition for an IP filter rule. The IpTimeCondition statement specifies when an IP filter or a manual IPSec tunnel is active. You can prescribe not only what traffic is allowed, but when that traffic is allowed, in a way that is not disruptive and without having to edit the IP security policy files.

For more details about the IpTimeCondition statement, see z/OS Communications Server: IP Configuration Reference.

Special considerations when using IP security for IPv6

This topic describes considerations for using IP security for IPv6.

Neighbor discovery and multicast listener discovery

For IPv6, TCP/IP uses the neighbor discovery protocol, which provides the following functions:

- Address resolution (neighbor solicitations and neighbor advertisements to perform ARP functions for IPv6)
- Duplicate address detection (neighbor solicitations and neighbor advertisements to ensure unique IP addresses)
- Router discovery (router solicitations and router advertisements to keep track of neighboring routers)
- Neighbor unreachability detection (to keep track of the reachability of neighbors)

TCP/IP also uses multicast listener discovery (MLD), which notifies routers that nodes are ready to receive multicast packets.

Neighbor discovery and MLD are implemented using ICMPv6 packets.

When the stack is enabled for IP security for IPv6, TCP/IP performs IP filtering for these packets and denies these packets by default. You must consider these packets when configuring filter rules for IPv6. For example, if you configure permit rules for IPv6 TCP or UDP traffic over an OSA-Express QDIO interface, but do not configure permit rules for the neighbor discovery address resolution flows over the interface, the traffic will not succeed because the stack denies the ICMPv6 neighbor solicitation and neighbor advertisement packets during address resolution. Also, you must configure permit filter rules for neighbor solicitation and neighbor advertisement packets, so that the stack can perform duplicate address detection (and learn about valid duplicate addresses).

In addition, for the stack to be able to learn about neighbors, you must configure permit filter rules for outbound router solicitations and inbound router advertisements. Similarly, for the stack to be able to perform the MLD listener function, you must configure permit filter rules for outbound MLD listener reports and inbound MLD listener query packets.

For example filter rules to permit neighbor discovery and MLD packets, see the TCP/IP sample profile and policy sample.

Guideline: Configure permit rules for neighbor discovery and MLD packets.

Result: z/OS Communications Server does not apply IPSec protection for outbound neighbor discovery or MLD packets, but does apply permit and deny actions for these packets. If such an outbound packet matches an IP filter rule that specifies an IPsec action, the stack permits the packet but does not encapsulate it.

Stateless address autoconfiguration

If you use autoconfiguration, your IPv6 addresses might not be predictable. To configure IP filter rules for dynamic Security Associations with autoconfigured IPv6 addresses, you need to specify the IP addresses using wildcards.

Manual Security Associations typically use specific IP addresses for the endpoints. You can use wildcards for the security endpoint addresses so that the data endpoints and security endpoints are considered identical. Alternatively, you can use predictable IPv6 addresses for the security endpoints. You can obtain predictable IPv6 addresses by configuring full 128-bit IPv6 addresses on your INTERFACE statements, by specifying the INTFID keyword on your INTERFACE statements, or by using VIPAs.

IPv6-specific protocols

The protocol number for ICMPv6 (58) is different from the protocol number for IPv4 ICMP (1); remember this when configuring IPv6 filter rules.

The handling of fragments in IPv6 is different from IPv4. For information about fragmented IPv6 packets, see “Routed traffic and fragmented packets” on page 897.

IPv6 address types

You can configure IP filter rules for any valid IPv6 address, including multicast addresses, link-local addresses, IPv4-mapped addresses, and so on.

You can configure IPSec tunnels for link-local and global addresses. For multicast addresses, you can configure manual tunnels but not dynamic tunnels. You cannot configure IPSec tunnels for IPv4-mapped addresses.

IPv6 extension headers

If an IPv6 packet contains an extension header that is a type 0 or type 2 routing header, the stack performs IP filtering using the final destination address of the packet based on the routing header contents, rather than using the destination address in the IP header.

Considerations for IPv6 OSPF security

IPv4 OSPF authentication is implemented within the IPv4 OSPF protocol. However, IPv6 OSPF security (both authentication and encryption) is implemented by using IPSec. Because OSPF uses both multicast messages and unicast messages, it is not possible to use dynamic tunnels for OSPF traffic. Instead, manual tunnels must be used. The IBM Configuration Assistant for z/OS Communications Server automates the process of creating IPv6 OSPF tunnels. The following information describes the process of manually creating the IPv6 OSPF tunnel definitions.

It is expected that the same manual tunnel is to be used for all link-local unicast and multicast traffic. Additional tunnels might be used for IPv6 OSPF virtual links.

Because multicast traffic is one-to-many, the manual tunnel must use the same Security Parameter Index (SPI) and keys for inbound and outbound traffic. Whatever SPI values and keys are used must be coordinated with all IPv6 OSPF peers on the LAN segment. Also, because this manual tunnel is to be used to protect traffic with various source and destination addresses, you must specify any6 for the local

and remote security endpoint locations. The following example uses AH authentication using the SHA algorithm, and ESP encryption using the DES algorithm.

```
IpManVpnAction tunnel-ipv6ospf-internal
{
  LocalSecurityEndpointAddr  any6
  RemoteSecurityEndpointAddr any6
  HowToAuth                  AH HMAC_SHA1
    AuthOutboundSa      2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
    AuthInboundSa       2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
  HowToEncrypt               DES
    EncryptOutboundSa   2701 0x3e6dcf72459ef551
    EncryptInboundSa    2701 0x3e6dcf72459ef551
  HowToEncap                 transport
}
```

For the filter rules, you first need to create an IP service to describe the OSPF traffic. To distinguish the traffic, you specify the OSPF protocol, and the SECCLASS of the interfaces on which the traffic will flow. For the purpose of this example, assume that the interfaces for the LAN segment that is being protected are defined with SECCLASS 10.

```
IpService service-ipv6ospf-internal
{
  Protocol      ospf
  Direction     bidirectional
  Routing       local
  SecurityClass 10
}
```

You now need to define three filter rules to match the OSPF traffic. The first filter rule matches all link-local unicast traffic on the LAN segment:

```
IpFilterRule ipv6ospf-unicast-internal
{
  IpSourceAddr             fe80::/10
  IpDestAddr               fe80::/10
  IpServiceRef             service-ipv6ospf-internal
  IpGenericFilterActionRef ipsec-nolog
  IpManVpnActionRef        tunnel-ipv6ospf-internal
}
```

The remaining two filter rules are for the OSPF link-local multicast traffic. The first rule matches outbound multicast traffic, which has a link-local unicast source address and a link-local multicast destination address. The second rule matches inbound multicast traffic, which has a remote (destination) address that is link-local unicast, and a local (source) address that is link-local multicast. These rules are as follows:

```
IpFilterRule ipv6ospf-outbound-multicast-internal
{
  IpSourceAddr             fe80::/10
  IpDestAddr               ff02::/16
  IpServiceRef             service-ipv6ospf-internal
  IpGenericFilterActionRef ipsec-nolog
  IpManVpnActionRef        tunnel-ipv6ospf-internal
}
IpFilterRule ipv6ospf-inbound-multicast-internal
{
  IpSourceAddr             ff02::/16
  IpDestAddr               fe80::/10
  IpServiceRef             service-ipv6ospf-internal
  IpGenericFilterActionRef ipsec-nolog
  IpManVpnActionRef        tunnel-ipv6ospf-internal
}
```

Virtual links

You can also configure IPsec protection for IPv6 OSPF virtual links. You can use the following method to configure manual tunnels for IPv6 OSPF virtual links.

Guideline: Because virtual links use global unicast addresses, you should use dynamic tunnels for IPv6 OSPF virtual links whenever possible. Dynamic tunnels provide numerous benefits over manual tunnels.

Because the virtual link addresses are not known beforehand, you must specify the security endpoint addresses for the manual tunnel with a wildcard value. The Security Association does not protect multicast traffic, so the inbound and outbound SPI values and keys do not need to be identical. Whatever SPI values and keys you use must be coordinated with the virtual link peer.

```
IpManVpnAction tunnel-ipv6ospfvirt-internal
{
  LocalSecurityEndpointAddr  any6
  RemoteSecurityEndpointAddr any6
  HowToAuth                  AH HMAC_SHA1
    AuthOutboundSa      2702 0xf5c58a6e6f0761b68f424f39257f0ea89a4be3b4
    AuthInboundSa       2703 0x39a341ed1b7127b7905df2411ed0770854b54d10
  HowToEncrypt               DES
    EncryptOutboundSa   2704 0xb9571d20fe98ecca
    EncryptInboundSa    2705 0xfeba84c113fb40ed
  HowToEncap                 transport
}
```

Only one filter rule is needed for the virtual link. For this example, the same IP service used for the link-local traffic is used, which restricts this filter rule to OSPF traffic flowing over interfaces with SECCLASS 10. Because the virtual link addresses are not known beforehand, the filter rule addresses are specified with a wildcard value to include all global unicast addresses. If you have more specific information about the address prefixes for the virtual link endpoints, you can use this to further restrict the addresses on the filter rule.

```
IpFilterRule ipv6ospf-virtual-internal
{
  IpSourceAddr              4000::/3
  IpDestAddr                4000::/3
  IpServiceRef             service-ipv6ospf-internal
  IpGenericFilterActionRef ipsec-nolog
  IpManVpnActionRef        tunnel-ipv6ospfvirt-internal
}
```

Default IP filter policy and IP security policy

Policy Agent provides IP filter policy to the stack, as defined by the IP security policy configuration files. For the stack to be enabled for IP security for IPv4, the TCP/IP profile must have IPCONFIG IPSECURITY coded. For the stack to be enabled for IP security for IPv6, the TCP/IP profile must also have IPCONFIG6 IPSECURITY coded. For the stack to receive configured policy, the Policy Agent must be active. IP filter policy is installed when Policy Agent and the stack are active. If Policy Agent is not active when an IPSECURITY-enabled stack initializes, the stack cannot be provided with IP filter rules from that policy. Therefore, in the interest of network security, the stack provides a default IP filter policy when an IP filter policy is unavailable from Policy Agent. The default IP filter policy effectively denies all network traffic, with the exception of some select ICMP and ICMPv6 messages that are necessary for internal stack function. These deny rules are not explicitly coded, but rather are always implicitly added at any time that the default IP filter policy is in effect. The default IP filter policy can be in effect at times other than stack initialization. Default IP filter policy is in effect in all of the following cases:

- An IPSECURITY-enabled TCP/IP stack has been started but Policy Agent has not been started, or an IPSECURITY-enabled TCP/IP stack has been started but Policy Agent has not completely initialized.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, but no IP security configuration file exists.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, but the IP security configuration file contains errors.
- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, but no IpFilterPolicy statement exists in either IP security configuration file.

- An IPSECURITY-enabled TCP/IP stack and Policy Agent have been started, and IP filter policy has been provided to the stack, but the stack is using the default IP filter policy because the **ipsec -f default** command was issued.

This default behavior ensures that network security is not compromised in the event that IP filter policy is not installed, and is consistent with a secure default-deny policy. However, you can modify the default IP filter policy by coding an IPSECRULE statement (for IPv4) or IPSEC6RULE statement (for IPv6) in the TCP/IP profile. The IPSECRULE and IPSEC6RULE statements describe the attributes of the IP traffic that is allowed when the default policy is active. Because the default behavior is to deny all network traffic, IPSECRULE and IPSEC6RULE statements are always permit rules that denote exceptions to the default-deny policy.

It is also important to note that neither the default IP filter policy nor a modified default IP filter policy provides authentication and encryption capabilities such as those provided by a complete IP security policy; it offers the stack only the ability to perform simple IP filtering in the absence of an IP security policy.

Modifying the default IP filter policy

The default IP filter policy is in effect when the IP security policy, as configured in the Policy Agent, is not available. The default IP filter policy denies all network traffic, unless you modify the TCP/IP profile.

Two IP filters are created by the default IP filter policy, one denying IPv4 inbound traffic and the other denying IPv4 outbound traffic. If IP security for IPv6 is also enabled, the stack also creates two similar filters, one denying IPv6 inbound traffic and the other denying IPv6 outbound traffic. For example, assuming that the default IP filter policy is active, the following sample of the **ipsec -f display** command shows that SYSDEFAULTDENYRULE was added to the filter table:

```
ipsec -f display

CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 09:50:02 2012
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Profile   Scope:    Current            TotAvail: 8
Logging:  On              Predecap: Off                DVIPSec:  No
NatKeepAlive:  0          FIPS140:  No
Defensive Mode: Inactive

FilterName:                   SYSDEFAULTRULE.1
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
```

```
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 09:44:37
UpdateTime:                     2012/02/14 09:44:37
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  49
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
************************************************************************
FilterName:                     SYSDEFAULTRULE.1
FilterNameExtension:            2
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Inbound
OnDemand:                       n/a
SecurityClass:                  0
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 09:44:37
UpdateTime:                     2012/02/14 09:44:37
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  94
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
************************************************************************
FilterName:                     SYSDEFAULTRULE.2
FilterNameExtension:            1
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
```

```
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                5
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTRULE.2
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
```

```
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                4
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
```

```
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Deny
Scope:                         Both
Direction:                     Inbound
OnDemand:                      n/a
SecurityClass:                 0
Logging:                       None
LogLimit:                      n/a
Protocol:                      All
ICMPType:                      n/a
ICMPTypeGranularity:           n/a
ICMPCode:                      n/a
ICMPCodeGranularity:           n/a
OSPFType:                      n/a
TCPQualifier:                  n/a
ProtocolGranularity:           n/a
SourceAddress:                 0.0.0.0
SourceAddressPrefix:           0
SourceAddressRange:            n/a
SourceAddressGranularity:      n/a
SourcePort:                    n/a
SourcePortRange:               n/a
SourcePortGranularity:         n/a
DestAddress:                   0.0.0.0
DestAddressPrefix:             0
DestAddressRange:              n/a
DestAddressGranularity:        n/a
DestPort:                      n/a
DestPortRange:                 n/a
DestPortGranularity:           n/a
OrigRmtConnPort:               n/a
RmtIDPayload:                  n/a
RmtUdpEncapPort:               n/a
CreateTime:                    2012/02/14 09:44:37
UpdateTime:                    2012/02/14 09:44:37
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 0
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
************************************************************************
FilterName:                    SYSDEFAULTDENYRULE
FilterNameExtension:           3
GroupName:                     n/a
LocalStartActionName:          n/a
VpnActionName:                 n/a
TunnelID:                      0x00
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Deny
Scope:                         Both
Direction:                     Outbound
OnDemand:                      n/a
SecurityClass:                 0
Logging:                       None
LogLimit:                      n/a
Protocol:                      All
ICMPType:                      n/a
ICMPTypeGranularity:           n/a
ICMPCode:                      n/a
ICMPCodeGranularity:           n/a
OSPFType:                      n/a
TCPQualifier:                  n/a
ProtocolGranularity:           n/a
SourceAddress:                 ::
SourceAddressPrefix:           0
SourceAddressRange:            n/a
SourceAddressGranularity:      n/a
SourcePort:                    n/a
SourcePortRange:               n/a
SourcePortGranularity:         n/a
DestAddress:                   ::
DestAddressPrefix:             0
```

```
DestAddressRange:              n/a
DestAddressGranularity:        n/a
DestPort:                      n/a
DestPortRange:                 n/a
DestPortGranularity:           n/a
OrigRmtConnPort:               n/a
RmtIDPayload:                  n/a
RmtUdpEncapPort:               n/a
CreateTime:                    2012/02/14 09:44:37
UpdateTime:                    2012/02/14 09:44:37
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 0
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
*******************************************************************************
FilterName:                    SYSDEFAULTDENYRULE
FilterNameExtension:           4
GroupName:                     n/a
LocalStartActionName:          n/a
VpnActionName:                 n/a
TunnelID:                      0x00
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Deny
Scope:                         Both
Direction:                     Inbound
OnDemand:                      n/a
SecurityClass:                 0
Logging:                       None
LogLimit:                      n/a
Protocol:                      All
ICMPType:                      n/a
ICMPTypeGranularity:           n/a
ICMPCode:                      n/a
ICMPCodeGranularity:           n/a
OSPFType:                      n/a
TCPQualifier:                  n/a
ProtocolGranularity:           n/a
SourceAddress:                 ::
SourceAddressPrefix:           0
SourceAddressRange:            n/a
SourceAddressGranularity:      n/a
SourcePort:                    n/a
SourcePortRange:               n/a
SourcePortGranularity:         n/a
DestAddress:                   ::
DestAddressPrefix:             0
DestAddressRange:              n/a
DestAddressGranularity:        n/a
DestPort:                      n/a
DestPortRange:                 n/a
DestPortGranularity:           n/a
OrigRmtConnPort:               n/a
RmtIDPayload:                  n/a
RmtUdpEncapPort:               n/a
CreateTime:                    2012/02/14 09:44:37
UpdateTime:                    2012/02/14 09:44:37
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 0
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
*******************************************************************************

8 entries selected
```

Use IPSECRULE and IPSEC6RULE for permit rules that denote exceptions to the default-deny policy. When the default IP filter policy is active, these permit rules appear in the default IP filter table before the SYSDEFAULTDENYRULE entries. Typically, these exceptions are few and are used for administrative access to the system when IP security policy is unavailable. For instance, a sample default set of permit rules might include entries to provide the following access:

- Administrative access
- Basic network services, such as DNS and OSPF routing advertisements
- Use of ping to test for server availability

IPSECRULE and IPSEC6RULE entries are coded in the IPSEC block of the TCP/IP profile. They describe the attributes of the IP traffic that is allowed when the default IP filter policy is active. These rules can specify criteria such as source address, destination address, protocol, source port, destination port, routing, direction, and security class.

IPSECRULE and IPSEC6RULE entries always have an action of permit; there is no action specification for deny or permit with IPSec protection.

Assuming the administrative machine has an IP address of 9.1.1.2 and connects through an interface whose security class is 100, the following sample IPSECRULE entries enable the z/OS system to communicate IPv4 DNS queries and OSPF routing advertisements to anyone, while giving blanket access to the administrator. The example IPSEC6RULE entry permits any ICMPv6 packets. An asterisk (*) is the default and represents all, indicating that any packet matches this attribute.

```
IPSEC LOGENable
; Rule      SrcAddr DstAddr   Logging Protocol   SrcPort    DestPort     Routing  Secclass

; OSPF protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO OSPF

; IGMP protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO 2

; DNS queries to UDP port 53
  IPSECRule *       *         NOLOG   PROTO UDP  SRCPort *  DESTport 53

; Administrative access
  IPSECRule *       9.1.1.2   LOG                                          SECCLASS 100

; ICMPv6 protocol
  IPSEC6Rule *      *         NOLOG   PROTO ICMPv6
ENDIPSEC
```

Any IPSECRULE and IPSEC6RULE entries that are coded are given a system-generated name for purposes of display. These rules are prefixed with the string SYSDEFAULTRULE. The IPSECRULE entries that are configured in the previous example are reflected in the **ipsec -f display** command as shown in the following example. Two rules are created for each IPSECRULE or IPSEC6RULE entry configured, one for each direction. By default, IPSECRULE and IPSEC6RULE entries define bidirectional filters. If you want to limit the filter to inbound or outbound traffic, use the DIRECTION parameter to specify direction of the traffic to be filtered. Because the five rules that are configured in the example IPSEC block are bidirectional, they have been expanded into 10 IP filters. Four additional filters have been created for the default-deny behavior. Each rule is given a unique filter name comprised of SYSDEFAULTRULE.*number*, where *number* is a numerical extension that indicates the relative order of the rule in the IP filter table. Because each bidirectional IPSECRULE and IPSEC6RULE statement results in multiple filter rules with the same name (inbound and outbound), a FilterNameExtension value is assigned by the system to uniquely identify each filter.

```
ipsec -f display

CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 09:55:31 2012
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Profile   Scope:    Current           TotAvail: 14
Logging:  On              Predecap: Off               DVIPSec:  No
NatKeepAlive:  0          FIPS140:  No
Defensive Mode: Inactive

FilterName:                   SYSDEFAULTRULE.1
```

```
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     OSPF(89)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     All
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:55:04
UpdateTime:                   2012/02/14 09:55:04
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                3
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTRULE.1
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     OSPF(89)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     All
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
```

```
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 09:55:04
UpdateTime:                     2012/02/14 09:55:04
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  4
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
************************************************************************
FilterName:                     SYSDEFAULTRULE.2
FilterNameExtension:            1
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Outbound
OnDemand:                       n/a
SecurityClass:                  0
Logging:                        None
LogLimit:                       n/a
Protocol:                       IGMP(2)
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 09:55:04
UpdateTime:                     2012/02/14 09:55:04
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
************************************************************************
```

```
FilterName:                    SYSDEFAULTRULE.2
FilterNameExtension:           2
GroupName:                     n/a
LocalStartActionName:          n/a
VpnActionName:                 n/a
TunnelID:                      0x00
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Permit
Scope:                         Local
Direction:                     Inbound
OnDemand:                      n/a
SecurityClass:                 0
Logging:                       None
LogLimit:                      n/a
Protocol:                      IGMP(2)
ICMPType:                      n/a
ICMPTypeGranularity:           n/a
ICMPCode:                      n/a
ICMPCodeGranularity:           n/a
OSPFType:                      n/a
TCPQualifier:                  n/a
ProtocolGranularity:           n/a
SourceAddress:                 0.0.0.0
SourceAddressPrefix:           0
SourceAddressRange:            n/a
SourceAddressGranularity:      n/a
SourcePort:                    n/a
SourcePortRange:               n/a
SourcePortGranularity:         n/a
DestAddress:                   0.0.0.0
DestAddressPrefix:             0
DestAddressRange:              n/a
DestAddressGranularity:        n/a
DestPort:                      n/a
DestPortRange:                 n/a
DestPortGranularity:           n/a
OrigRmtConnPort:               n/a
RmtIDPayload:                  n/a
RmtUdpEncapPort:               n/a
CreateTime:                    2012/02/14 09:55:04
UpdateTime:                    2012/02/14 09:55:04
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 0
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
*************************************************************************
FilterName:                    SYSDEFAULTRULE.3
FilterNameExtension:           1
GroupName:                     n/a
LocalStartActionName:          n/a
VpnActionName:                 n/a
TunnelID:                      0x00
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Permit
Scope:                         Local
Direction:                     Outbound
OnDemand:                      n/a
SecurityClass:                 0
Logging:                       None
LogLimit:                      n/a
Protocol:                      UDP(17)
ICMPType:                      n/a
ICMPTypeGranularity:           n/a
ICMPCode:                      n/a
ICMPCodeGranularity:           n/a
OSPFType:                      n/a
TCPQualifier:                  n/a
ProtocolGranularity:           n/a
SourceAddress:                 0.0.0.0
SourceAddressPrefix:           0
SourceAddressRange:            n/a
```

```
SourceAddressGranularity:     n/a
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     53
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:55:04
UpdateTime:                   2012/02/14 09:55:04
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTRULE.3
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     UDP(17)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   53
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:55:04
UpdateTime:                   2012/02/14 09:55:04
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
```

```
*******************************************************************************
FilterName:                     SYSDEFAULTRULE.4
FilterNameExtension:            1
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Outbound
OnDemand:                       n/a
SecurityClass:                  100
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    9.1.1.2
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 09:55:04
UpdateTime:                     2012/02/14 09:55:04
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*******************************************************************************
FilterName:                     SYSDEFAULTRULE.4
FilterNameExtension:            2
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Inbound
OnDemand:                       n/a
SecurityClass:                  100
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  9.1.1.2
SourceAddressPrefix:            n/a
```

```
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:55:04
UpdateTime:                   2012/02/14 09:55:04
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*******************************************************************************
FilterName:                   SYSDEFAULTRULE.5
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     ICMPV6(58)
ICMPType:                     All
ICMPTypeGranularity:          n/a
ICMPCode:                     All
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:55:04
UpdateTime:                   2012/02/14 09:55:04
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
```

```
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTRULE.5
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     ICMPV6(58)
ICMPType:                     All
ICMPTypeGranularity:          n/a
ICMPCode:                     All
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:55:04
UpdateTime:                   2012/02/14 09:55:04
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
```

```
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*******************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                0.0.0.0
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  0.0.0.0
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                1
```

```
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*******************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          3
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*******************************************************************************
FilterName:                   SYSDEFAULTDENYRULE
FilterNameExtension:          4
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
```

```
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 09:44:37
UpdateTime:                   2012/02/14 09:44:37
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*********************************************************************

14 entries selected
```

For an IP security-enabled stack, one of the two security policies is always in effect, either the default policy or the IP security policy as defined in Policy Agent. The policy that is in effect at any given time is considered to be the active policy. The source field of the report header for the **ipsec -f display** command can be used to determine which policy is currently active as follows:

```
ipsec -f display

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 09:55:31 2010
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Profile   Scope:    Current            TotAvail: 14
Logging:  On              Predecap: Off                DVIPSec:  No
NatKeepAlive:  0          FIPS140:  No
Defensive Mode: Inactive

FilterName:                   SYSDEFAULTRULE.1
FilterNameExtension:          1
```

`Stack Profile` indicates that the default IP filter policy is active; `Stack Policy` would indicate that the IP security policy as defined in Policy Agent is active.

You can also choose which policy is the active policy. The **ipsec** command provides the ability to switch the active policy between the default IP filter policy in the TCP/IP profile and the IP security policy in Policy Agent. Issuing the **ipsec -f default** command causes the default policy to become the active policy, while issuing the **ipsec -f reload** command reloads the IP security policy from Policy Agent, provided that Policy Agent is active and that IP security has been correctly configured.

There are cases in which you might want to switch to the default policy. For instance, in the event of a security breach, the **ipsec -f default** command allows only the network traffic that has been explicitly coded in the TCP/IP profile IPSEC block, which typically permits only administrative access.

It is important to use the **ipsec -f default** command with discretion. Issuing **ipsec -f default** on an operational system can have a dramatic impact and cause packets to be dropped, depending on how the IPSECRULE and IPSEC6RULE entries are coded. Consider the use of **ipsec -f default** as equivalent to a shutdown, and do not use it in normal circumstances unless the following conditions are true:

- The system is in maintenance mode.
- There is no productive work being done on the system.

- The site policy for IP traffic is default-allow and the appropriate IPSECRULE and IPSEC6RULE entries were coded.

Malicious use of network resources can be identified by unusual IP traffic conditions, console messages, or log inspection. Should a security concern arise in a production environment, the following steps should be performed quickly to minimize the impact of potentially restricted flow of IP traffic.

1. To secure the system, switch to the default IP filter policy by issuing the **ipsec -f default** command.
2. Analyze system logs to determine the nature and the source of the security concern. If packet logging was active for the time period in question, then inspect the TRMD packet filtering log entries.
3. Update the IP security policy configuration to alleviate the security concern. For example, add an IpFilterRule that denies any suspicious IP address or port use, and activate IP filter logging.
4. Activate the updated IP security policy from the Policy Agent by issuing the **ipsec -f reload** command.
5. Verify that the security concern is eliminated. Monitor the network traffic and inspect the system logs, including any related IP filter log messages.

IP filter logging

Monitoring network events is an important aspect of network security. Logs can be used to verify that policies have been correctly configured and enforced, or to gather statistics on any traffic of interest. For instance, traffic that is persistently denied might be suspect. With IP filter logging, you can inspect any traffic on the system and even fine tune the configuration to show only those entries of interest. Logging can be controlled at the global level or at the individual rule level, including the ability to specify whether to log permitted traffic, denied traffic, or both. The IP filter log entries provide detailed information about each packet, including the rule that the packet matched and any pertinent IPSec information.

TRMD and syslogd provide the logging service for IP security. If running in the common INET environment, you must configure one instance of TRMD for each stack on a z/OS system.

Guideline: Exhaustive logging of IP traffic can have a negative effect on performance. If logging is excessive, it can be turned off temporarily at the global level while the appropriate logging modifications are made to the individual IP filter rules. IP filter logging is controlled by the IpFilterLogging parameter on the IpGenericFilterAction statement. For more details, see z/OS Communications Server: IP Configuration Reference.

IP filter discard action

When packets are denied by IP filter policy, IP security supports sending an ICMP or ICMPv6 destination unreachable message, which indicates that a packet is administratively prohibited. You can enable this action for the implicit Policy Agent deny filter rules using the ImplicitDiscardAction parameter on the IpFilterPolicy statement, and for individual filter rules using the DiscardAction parameter on the IpGenericFilterAction statement.

Rule: To make a system effectively invisible to attackers, choose the discard action Silent. Choose the discard action ICMP only in cases in which doing so is a useful diagnostic aid, such as for users of internal networks.

Data encryption and authentication - IPSec

To participate in a virtual private network (VPN), a host must encrypt and authenticate individual IP packets between itself and another communicating host. IPSec is one of several mechanisms for achieving this, and one of the more versatile.

IPSec is defined by the IPSec working group of the IETF. It provides authentication, integrity, and data privacy between any two IP entities. Management of cryptographic keys and Security Associations can be either manual or dynamic using an IETF-defined key management protocol called Internet Key Exchange (IKE).

IPSec provides flexible building blocks that can support a variety of configurations. Because an IPSec Security Association can exist between any two IP entities, it can protect a segment of the path or the entire path. The main advantage of using IPSec for data encryption and authentication is that IPSec is implemented at the IP layer. Consequently, any network traffic that is carried by an IP network is eligible to use IPSec services without any special changes to higher level protocols that are used by applications. However, if the system is using any of these alternate security protocols to secure specific applications, IP filtering can be used to avoid the overhead of multiple security protocols. For example, you might want to exclude web traffic (based on the well-known secure port of the web server, port 443) from IPSec coverage because you would like to use SSL.

IPSec enables the creation of VPNs. A VPN enables an enterprise to extend its network across a public network, such as the Internet, through a secure tunnel using Security Associations. IPSec VPNs enable the secure transfer of data over the public Internet for same-business and business-to-business communications, and protect sensitive data within an enterprise's internal network.

IPSec uses IP filtering to determine which traffic should be protected by IPSec. A special type of filter action specifies to permit the traffic, but only with IPSec protection. The IP filters represent IP security policy to the stack by specifying the traffic that requires IPSec protection. The filters are also used in locating the outbound IPSec Security Association, and for verifying that inbound traffic is received using the correct Security Association.

The IETF has standardized the IPSec protocol suite and key management schemes in a series of IPSec RFCs. For more information on these RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

IPSec has three major components:

- IP Authentication Header (AH)
- IP Encapsulating Security Payload (ESP)
- Internet Key Exchange (IKE)

AH and ESP protocols

IPSec uses two distinct protocols, Authentication Header (AH) and Encapsulating Security Payload (ESP), which are defined by the IETF.

The AH protocol provides a mechanism for authentication only. AH provides data integrity, data origin authentication, and an optional replay protection service. Data integrity is ensured by using a message digest that is generated by an algorithm such as HMAC-MD5 or HMAC-SHA. Data origin authentication is ensured by using a shared secret key to create the message digest. Replay protection is provided by using a sequence number field with the AH header. AH authenticates IP headers and their payloads, with the exception of certain header fields that can be legitimately changed in transit, such as the Time To Live (TTL) field.

The ESP protocol provides data confidentiality (encryption) and authentication (data integrity, data origin authentication, and replay protection). ESP can be used with confidentiality only, authentication only, or both confidentiality and authentication. When ESP provides authentication functions, it uses the same algorithms as AH, but the coverage is different. AH-style authentication authenticates the entire IP packet, including the outer IP header, while the ESP authentication mechanism authenticates only the IP datagram portion of the IP packet.

Either protocol can be used alone to protect an IP packet, or both protocols can be applied together to the same IP packet. The choice of IPSec protocol is determined by the security needs of your installation, and is configured by the administrator. It does not have to be applied system-wide, and can be configured differently for each set of connection endpoints. For a dynamic tunnel, the choice of IPSec protocol is configured using the IpDataOffer statement in an IP security policy configuration file. For a manual tunnel, the choice of IPSec protocol is configured using the IpManVpnAction statement in an IP security policy configuration file. For more details about the IpDataOffer statement and the IpManVpnAction statement, see z/OS Communications Server: IP Configuration Reference.

z/OS IP security requires authentication due to potential security exposures when encryption is used alone. Authentication can be provided by the ESP or AH protocol. The complete list of combinations for authentication and encryption that are provided by z/OS IP security and that can be used for a specific connection are shown in Table 49 on page 921.

Table 49. Possible authentication and encryption combinations for a connection

Encryption Protocol	Encryption Algorithm	Authentication Protocol	Authentication Algorithm
None	None.	ESP or AH	Any of the following algorithms: • HMAC_MD5 • HMAC_SHA1 • AES128_XCBC_96 • HMAC_SHA2_256_128 • HMAC_SHA2_384_192 • HMAC_SHA2_512_256 • AES_GMAC_128 • AES_GMAC_256
ESP	Any of the following algorithms: • DES • 3DES • AES_CBC KeyLength 128 • AES_CBC KeyLength 256	ESP or AH	Any of the following algorithms: • HMAC_MD5 • HMAC_SHA1 • AES128_XCBC_96 • HMAC_SHA2_256_128 • HMAC_SHA2_384_192 • HMAC_SHA2_512_256
ESP	Any of the following algorithms: • AES_GCM_16 KeyLength 128 • AES_GCM_16 KeyLength 256	ESP	NULL (AES_GCM provides built-in authentication)

Guideline: RFC 4835 discourages the use of DES. Use the AES encryption algorithms wherever possible for better security and interoperability.

Restriction: The combination of ESP protocol for encryption and AH protocol for authentication is not supported by IKEv2. If you are using IKEv2 and require both encryption and authentication, you should use ESP for both.

Encapsulation

In the process of applying either AH or ESP to an IP packet, the original IP packet is modified. Outbound packets are rebuilt with additional IPSec headers in a process known as encapsulation, while inbound packets are stripped of their IPSec headers in a process known as decapsulation. Before leaving a host, outbound packets are encapsulated using a cryptographic key that is known to both communicating hosts. Inbound packets are decapsulated on the receiving side using the same cryptographic key, thereby recovering the original datagram. If encryption is used, any packet that is intercepted on the IP network is unreadable to anyone without the encryption key. Any modifications to the IP packet while in transit are detected by authentication processing at the receiving host and discarded.

Transport mode and tunnel mode

The manner in which the original IP packet is modified depends on the encapsulation mode used. There are two encapsulation modes used by AH and ESP, transport and tunnel.

Transport mode encapsulation retains the original IP header. Therefore, when transport mode is used, the IP header reflects the original source and destination of the packet. Transport is most often used in a host-to-host scenario, where the data endpoints and the security endpoints are the same. A transport mode encapsulated datagram is routed, or transported, in the same manner as the original packet.

Figure 111 on page 922 shows an IPv4 packet that is encapsulated using AH in transport mode:

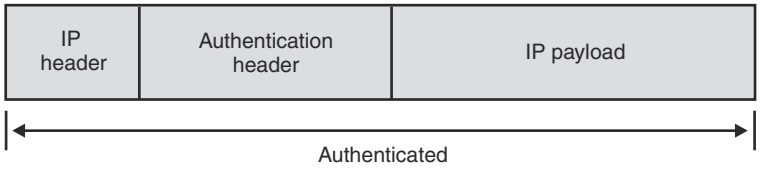


Figure 111. IPv4 packet encapsulated using AH in transport mode

Figure 112 on page 922 shows an IPv4 packet that is encapsulated using ESP in transport mode:

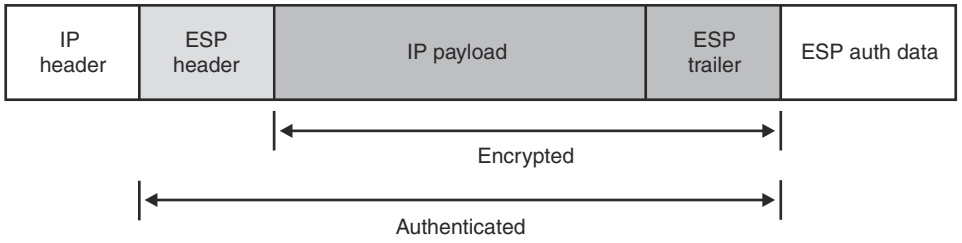


Figure 112. IPv4 packet encapsulated using ESP in transport mode

Figure 113 on page 922 shows an IPv6 packet that is encapsulated using AH in transport mode:

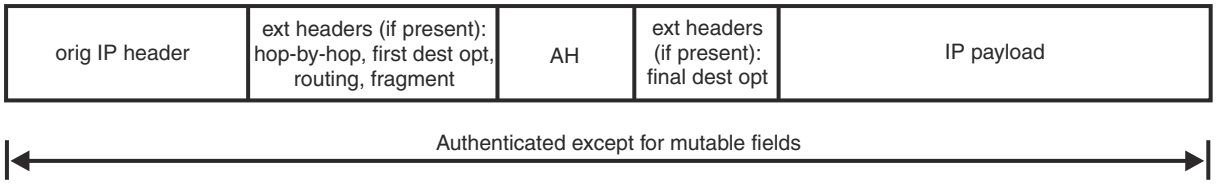


Figure 113. IPv6 packet encapsulated using AH in transport mode

For a description of the IPv6 mutable fields, see RFC 2402. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

Figure 114 on page 922 shows an IPv6 packet that is encapsulated using ESP in transport mode:

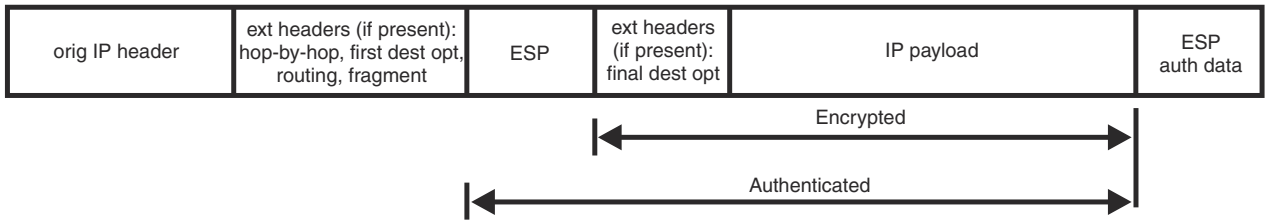


Figure 114. IPv6 packet encapsulated using ESP in transport mode

Tunnel mode encapsulation builds a new IP header containing the source and destination address of the security endpoints. When tunnel mode is used, the outer IP header reflects the source and destination of the security endpoints, which might or might not be the same as the original source and destination IP address of the data connection. The choice of transport or tunnel mode depends on the structure of the network and relies heavily on logical connections between the endpoints. Tunnel mode is required if

one of the IKE peers is a security gateway that is applying IPSec on behalf of another host or hosts. A datagram that is encapsulated in tunnel mode is routed, or tunneled, through the security gateways, with the possibility that the secure IPSec packet will not flow through the same network path as the original datagram. To successfully encapsulate and send an outbound packet, the route table must contain a route that can be used to reach the security gateway, as well as a route that can be used to reach the data endpoint. If policy-based routing is being used on a TCP/IP stack where IP security is active, it is important to understand how the two functions interact. For more information, see "Considerations for using policy-based routing with IP security" on page 356.

Figure 115 on page 923 shows an IPv4 packet that is encapsulated using AH in tunnel mode:

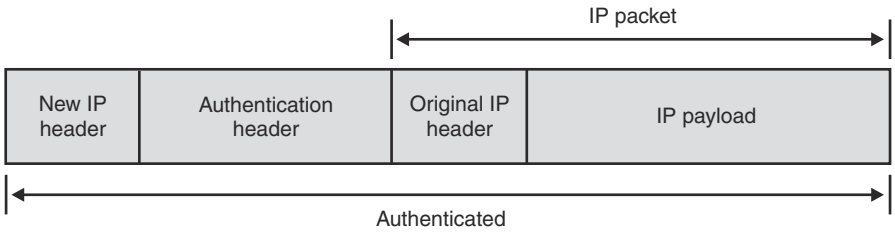


Figure 115. IPv4 packet encapsulated using AH in tunnel mode

Figure 116 on page 923 shows an IPv4 packet that is encapsulated using ESP in tunnel mode:

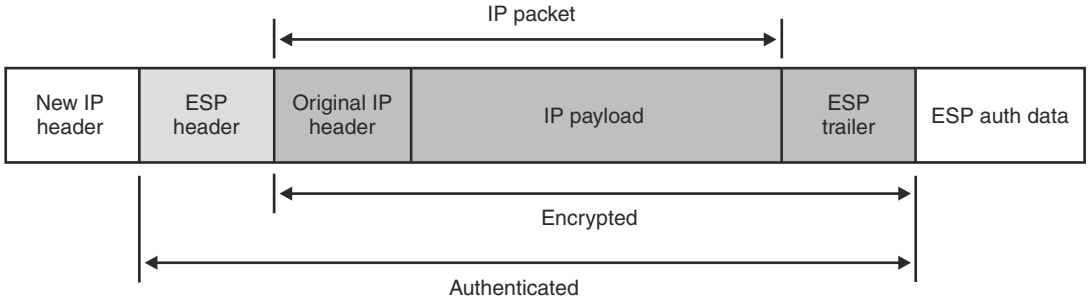


Figure 116. IPv4 packet encapsulated using ESP in tunnel mode

Figure 117 on page 923 shows an IPv6 packet that is encapsulated using AH in tunnel mode:

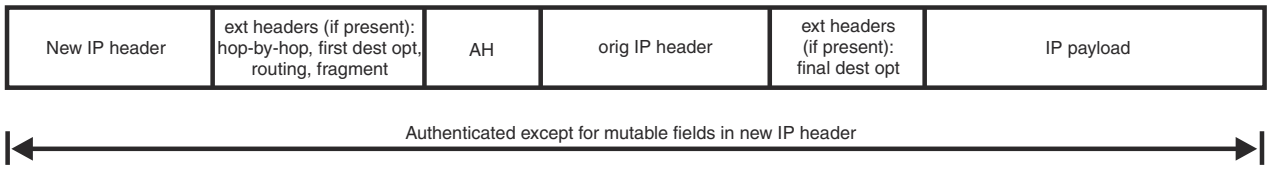


Figure 117. IPv6 packet encapsulated using AH in tunnel mode

For a description of the IPv6 mutable fields, see RFC 2402.

Figure 118 on page 923 shows an IPv6 packet that is encapsulated using ESP in tunnel mode:

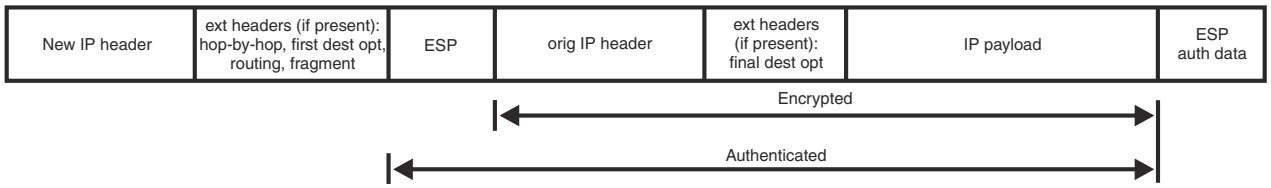


Figure 118. IPv6 packet encapsulated using ESP in tunnel mode

Do not confuse tunnel mode encapsulation with IKE tunnel or IPSec tunnel. In this context, tunnel refers only to the method by which IPSec packets are constructed, while IKE and IPSec tunnels are conceptually defined as secure logical connections between hosts. IPSec tunnels can use transport mode or tunnel mode encapsulation.

For a dynamic tunnel, the choice of encapsulation mode is configured using the IpDataOffer statement in an IP security policy configuration file. For a manual tunnel, the choice of IPSec protocol is configured using the IpManVpnAction statement in an IP security policy configuration file. For more details about the IpDataOffer statement and the IpManVpnAction statement, see z/OS Communications Server: IP Configuration Reference.

Guideline: In host-to-host scenarios, transport mode is commonly used because it does not incur the overhead of having to build an additional IP header. However, tunnel mode is equally valid.

Tip: Any configurations that include a secure gateway as an IKE endpoint require tunnel mode; however, if IKEv2 is used, then the selection of encapsulation mode is performed automatically. The IKED will use transport mode if it determines that the connection is host-to-host; otherwise the IKED will use tunnel mode encapsulation. This behavior can be overridden by using the HowToEncapIKEv2 parameter on the associated KeyExchangeAction statement; see the KeyExchangeAction statement in z/OS Communications Server: IP Configuration Reference for details.

Rule: If you are using Sysplex-Wide Security Associations, a dynamic Security Association cannot use a subnet or range that encompasses a DVIPA address. For takeover to work consistently, Security Associations must be negotiated for single IP addresses only. If two DVIPAs that are covered by the same Security Association are subsequently taken over by two different backup stacks, the coverage of the Security Association is ambiguous because it is then linked to two DVIPAs on two different stacks.

UDP encapsulation of IPSec ESP packets

When building an ESP packet, it can be further encapsulated by placing a UDP header in front of the ESP header. This is known as UDP encapsulation. UDP encapsulation is used to allow IPSec traffic to successfully traverse a NAT device. For more information on NAT traversal (NATT), see "IPSec and network address translation devices" on page 932. z/OS Communications Server supports NAT traversal for IPv4 traffic only.

z/OS Communications Server supports both tunnel and transport modes of UDP encapsulation.

As shown in Figure 119 on page 924, UDP-encapsulated transport mode inserts a UDP header in between the IP header and the ESP header of a normal transport mode ESP packet.

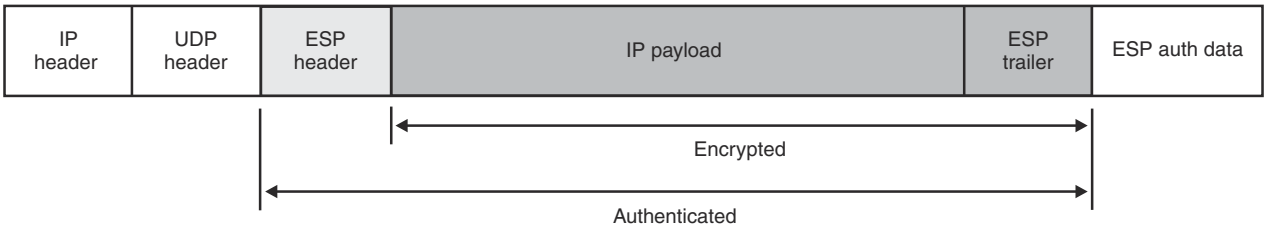


Figure 119. UDP-encapsulated transport mode

As shown in Figure 120 on page 924, UDP-encapsulated tunnel mode inserts a UDP header in between the new IP header and the ESP header of a normal tunnel mode ESP packet.

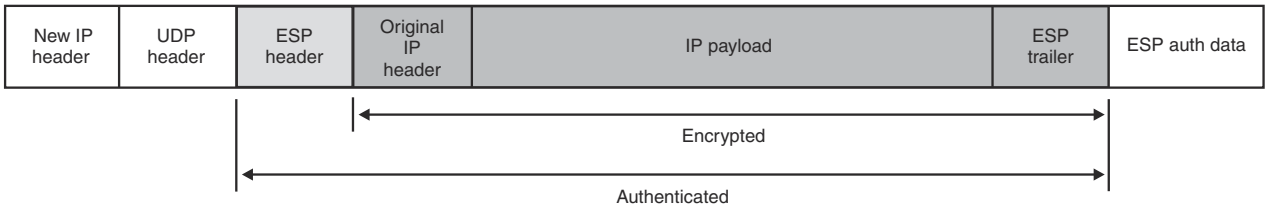


Figure 120. UDP-encapsulated tunnel mode

When an IPSec UDP-encapsulated packet is built, the source and destination port values in the UDP header are set to the IKE port value of 4500.

Configure the choice of transport or tunnel mode using the IpDataOffer statement in the IP security policy configuration file. For more details about the IpDataOffer statement, see z/OS Communications Server: IP Configuration Reference.

The decision to use a UDP-encapsulated mode is not configured, but instead inferred, when a NAT is detected between two IKE daemons.

IPSec and symmetric key management

At the center of encryption and authentication is the notion of a cryptographic key. Security endpoints use keys to encrypt and decrypt data. The IPSec protocols create Security Association keys that are directional. As shown in Figure 121 on page 925, the key that is used to encrypt outbound data on one host is used to decrypt the same data on the remote host, while the key that is used to encrypt data on the remote host is used to decrypt data on the local host.

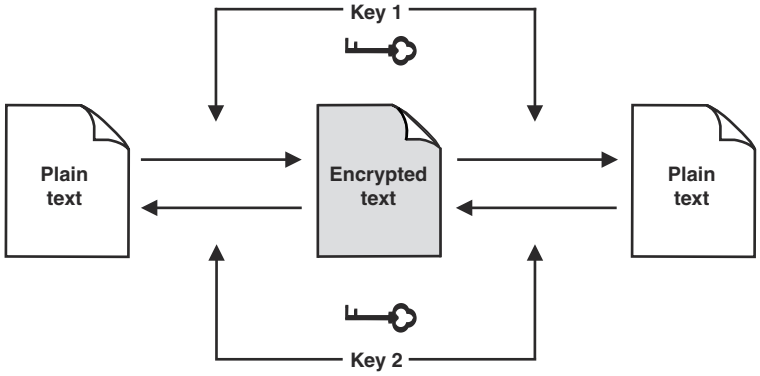


Figure 121. Symmetric encryption

This type of encryption is known as symmetric, because it requires that both hosts use the same keys on the same data.

Manual key management

IPSec keys and values can be configured manually. Manual management of keys was the only non-proprietary option for implementing IPSec before the standardization of the IKE protocols. In a manual IPSec configuration, the keys that are used to encrypt data, and the Security Parameter Index (SPI) values that are used to uniquely identify a Security Association, are determined by the administrator and configured beforehand on both hosts. However, the steps that you must take to manually generate encryption keys can become quite cumbersome, as shown in the following examples:

- A single IPSec connection requires two keys, because encryption keys are unidirectional (one for inbound traffic, one for outbound traffic).
- A single IPSec connection can require up to four SPI values, depending on the type of IPSec protection required.
- The keys and SPI values must be individually installed on both the local and remote system.
- To help ensure that the keys are not compromised, they need to be changed periodically and updated on both hosts. Manual IPSec has no key refresh capabilities unless the Security Associations are deactivated, reconfigured with the new key, and then reactivated.
- During the refresh maintenance period, IPSec functionality is unavailable, disrupting any IP traffic requiring IPSec protection.
- Because of the disruptive nature of key refresh with manual IPSec, key lifetimes are frequently defined with much larger values, thus increasing the security exposure.
- These steps must be carried out for each remote host that communicates with IPSec, with different keys for each of them.

In a small installation with minimal security concerns, this manual process can work quite well, and might actually be preferable. However, given the large number of communicating hosts in a typical

network, this manual process quickly becomes unwieldy, error prone, maintenance intensive, and not scalable. Nevertheless, z/OS IP security does support manual IPSec configuration for compatibility with systems that use it. If an IP filter rule specifies a manual ipsec action, the corresponding action is consulted to determine all aspects of the encryption and authentication policy for that data, including the identities of the communicating hosts, the keys that are used for encryption and authentication of outbound and inbound packets, and the SPI values. Manual IPSec protection is configured using the IpManVpnAction statement in an IP security policy configuration file. For more details about the IpManVpnAction statement, see z/OS Communications Server: IP Configuration Reference.

Dynamic key management - IKE and IPSec negotiations

The primary role of the IKE daemon in an IP security environment is the automatic management of cryptographic keys. Dynamic key management, as provided by the IKE daemon, removes much of the administrative burden that is associated with the creation, distribution, and maintenance of cryptographic keys. IKE provides the following services:

- Host authentication (ensuring that both hosts are certain of the other's identity)
- The negotiation of a Security Association as follows:
  - Agreeing on the type of traffic to be protected
  - Agreeing on the authentication and encryption algorithms to be used
  - Generating cryptographic keys
- Nondisruptive periodic refresh of keys
- The deletion of Security Associations whose lifetimes have expired

There are two versions of the IKE protocol: IKEv1 and IKEv2. The architectural differences between IKEv1 and IKEv2 are as follows:

- IKEv2 requires fewer network flows for Security Association establishment.
- IKEv2 provides a mechanism for re-keying an IKE Security Association without reauthentication, which reduces the CPU cost.
- IKEv2 allows each peer to manage its own values for lifetime and lifesize, whereas in IKEv1 Security Association lifetimes and lifesizes are negotiable attributes. This reduces coordination of configuration parameters between the peer nodes and avoids potential race conditions when the SAs expire.
- IKEv2 supports the hash and URL certificate encoding types, but IKEv1 does not.

The following difference between IKEv1 and IKEv2 is specific to the z/OS implementation:

- To use any authentication method for IKEv2 based on a digital signature, the certificate service must be provided by an NSS server. You cannot use the IKED native certificate service.

IKE operates at the application layer. IKE negotiations are communicated between two IKE peers by a series of UDP messages. Ports 500 and 4500 are used by the IKE daemon. For both IKEv1 and IKEv2, Security Association negotiation proceeds in phases. In each phase, IKE negotiates a Security Association with a remote host. During the initial phase, IKE negotiates a phase 1 Security Association, which establishes a secure channel over which the IKE peers communicate. The phase 2 Security Association is negotiated to provide data authentication and encryption for subsequent IP traffic. The distinction between a phase 1 Security Association and a phase 2 Security Association is what the Security Associations protect:

- Phase 1 Security Associations are used to protect IKE messages that are exchanged between two IKE peers, or security endpoints.
- Phase 2 Security Associations are used to protect IP traffic, as specified by the security policy for a specific type of traffic, between two data endpoints.

Every Security Association that the IKE daemon negotiates contains information about the type of traffic it is to protect, the IP addresses of the two endpoints, the type of encryption or authentication that is provided, the keys that are used to protect data, how often the keys are refreshed, and an identifier called a Security Parameter Index (SPI) that is used to uniquely identify the Security Association.

An IPSec configuration contains separate policies governing each phase. Although many of the same attributes apply for phase 1 and phase 2 negotiations and keys, there are two major differences:

- The phase 1 Security Association can specify only a single IP address for the security endpoints, while the phase 2 Security Association can specify a contiguous range or subnet as the data endpoint.
- The phase 1 Security Association must specify an encryption method, while encryption is optional for the phase 2 Security Association. An authentication method must be specified for both the phase 1 and phase 2 Security Association.

The exact specifications for each phase 1 Security Association and each phase 2 Security Association are configured in the IP security configuration file. The specific IP security policy statements that apply to the phase 1 and phase 2 specifications are the KeyExchangeOffer statement (phase 1) and the IpDataOffer statement (phase 2). For more details about the KeyExchangeOffer statement and the IpDataOffer statement, see z/OS Communications Server: IP Configuration Reference.

Because the IKE protocols deal with initializing keys, they must be capable of running over links where no security is assumed to exist. IKE addresses the problem of secure key distribution by automatically deriving the keying material using a Diffie-Hellman exchange during the IKE phase 1 negotiation. This automatic creation and distribution of the key during phase 1 eliminates the need to manually distribute the session key between remote sites. Besides the obvious administrative advantage of IKE, the manual method of key distribution is prone to key compromise.

In addition, IKE non-disruptively refreshes the session keys based on the security policy of the installation. IKE specifies that this can be based on time (lifetime) and bytes transmitted (life size). IKE provides a property called perfect forward secrecy (PFS), and if PFS is used, each phase 2 key is derived independently through a separate Diffie-Hellman exchange. With PFS, if a single key is compromised, the integrity of subsequently generated keys is not affected.

Phase 1

Before IKE can negotiate the security parameters and generate the keys that are used to protect data between the two hosts, it must have a way of protecting the negotiation itself. The IKE phase 1 negotiation provides this protection by performing two tasks:

- Authenticating the IKE peer

  Peer authentication is performed either by the pre-shared key method or a digital signature method. For details of peer authentication, see “Peer authentication” on page 928.
- Generating cryptographic keys

  A Diffie-Hellman exchange is performed to create a shared secret between the two IKE peers. This shared secret is then used in the generation of keying material. Keys to encrypt and authenticate messages sent during phase 2 are produced from this keying material. Cryptographic keys used by phase 2 Security Associations are generated from this keying material. The creation of the Diffie-Hellman shared secret is secure, but computationally expensive.

The phase 1 Security Association contains the following information:

- The key that is used to encrypt IKE messages
- The key that is used to authenticate IKE messages
- Keying material used to generate keys produced during phase 2
- The security endpoints (single IP addresses)
- The type of protection that is required (authentication and encryption)
- How often the keys should be renewed
- A Security Parameter Index (SPI) value, which is used together with the remote security endpoint IP address to uniquely identify the Security Association
- The Diffie-Hellman group, which is an attribute of the public key cryptography algorithm

Because the tasks of authentication and master key generation are so resource intensive, a phase 1 Security Association is usually refreshed less often than a phase 2 Security Association.

Peer authentication

Peer authentication is a critical part of a phase 1 negotiation. Before two hosts can participate in the negotiation of a Security Association, each must be authorized to negotiate with the other. IKE does not allow negotiation with a host that cannot be properly identified. IP addresses are not a guarantee of identity (an IP packet can be spoofed), and the IP address from an inbound packet alone is insufficient to prove the identity of a remote host. Therefore, the IKE daemon needs a more reliable method for determining the remote host's identity. This proof of identity is first presented during the phase 1 negotiation.

z/OS IP security implements two methods of host authentication as follows:

- Digital signature (RSA or ECDSA)
- Pre-shared key

Each of these authentication methods provides a way for hosts to verify the identity of the other, and while the pre-shared key mechanism is easier to configure, the digital signature methods are more versatile, secure, and scalable. The choice of authentication method is configured in an IP security policy configuration file for each IPSec connection, using the KeyExchangeOffer statement for IKEv1 and the KeyExchangeAction statement for IKEv2. For more details about the KeyExchangeOffer and KeyExchangeAction statements, see z/OS Communications Server: IP Configuration Reference.

Peer authentication is not the same as data authentication. Data authentication uses IPSec to authenticate an IP packet after an IPSec Security Association has been negotiated by two IKE daemons. Peer authentication is used during an IKE phase 1 negotiation to identify two IKE peers to each other before the establishment of any phase 2 Security Associations.

Guideline: As a matter of security, pre-shared keys should not be shared among multiple remote IKE peers. If the pre-shared key method of authentication is used, each remote host should have its own unique key and KeyExchangeRule specific to that host. If multiple remote hosts are identified by the same KeyExchangeRule, digital signature methods of peer authentication should be used.

Rule: For IKEv2, if digital signatures are used as the authentication method, then the IKED must be configured to use the certificate services of the NSSD. For more details about the configuring the IKED and the NSS IPSec client, see Chapter 18, "Network security services," on page 1083.

Identity information

Multiple identities can be assigned down to the level of individual IP addresses for each IKE negotiation, or for ease of configuration, a single identity can be assigned to represent the IKE daemon for the TCP/IP stack. With either method, the identity that is assigned must be an identity that is known to the remote IKE peer. Similarly, the local server must know the identity of any remote IKE peer with which it is to negotiate.

Guideline: Do not use the same identity on more than one TCP/IP stack or z/OS system image. IKED assumes that an identity is not used by any other stack, and this assumption can lead to a disruption of IPSec service for other stacks when identities are shared between stacks.

The identity can be one of the following types:

- X500dn (the host's X.500 distinguished name)
- IpAddr (an IP address)
- Fqdn (a fully qualified domain name)
- UserAtFqdn (an email address)
- KeyID (EBCDIC, ASCII or hexadecimal string)

  Restriction: An identity of KeyID is valid only when using preshared key authentication.

If a digital signature method of peer authentication is used, the local identity must be in an X.509 digital certificate on the IKE daemon's key ring for an IKEv1 negotiation, or on the NSS server's key ring for an IKEv1 or IKEv2 negotiation that uses the NSS certificate service. If the pre-shared key method of peer authentication is used, any of the identity types can be used to identify this IKE peer. Because digital

certificates are not used in the pre-shared key method, the only requirement for the use of an identity type is that it be known to both hosts.

Identity information is configured using the LocalSecurityEndpoint and RemoteSecurityEndpoint statements in an IP security policy configuration file. For more details about the LocalSecurityEndpoint and RemoteSecurityEndpoint statements, see z/OS Communications Server: IP Configuration Reference.

Digital signatures

Digital signature authentication methods include RSA and ECDSA. They use X.509 digital certificates. An X.509 digital certificate contains information that was verified by a certificate authority to uniquely identify a host. IKE peers exchange certificates during the phase 1 negotiation to identify each other. Of the information that is typically included in an X.509 digital certificate, there are four fields that are relevant to an IKE exchange:

Subject Name
A certificate's subject name is the unique name by which the host is known. The subject name is used to extract the host's X.500 distinguished name (X.500dn). IKE can use the X.500dn as an identifier for the remote host, or use one of the certificate's Subject Alternative Names.

Subject Alternative Name
An optional field, X.509 extensions, can contain a Subject Alternative Name field. Although the X.500dn is the most specific way to identify a certificate, the Subject Alternative Name can be used as well, often acting as a simple alias to identify the certificate. If the certificate includes optional X.509v3 extensions, the certificate can contain any or all of the following information:

- IpAddr - an IP address
- Fqdn - a fully qualified domain name
- UserAtFqdn - an email address

Subject Public Key Info
The public key is used for encrypting information that can be decrypted only with the certificate's private key. Likewise, information that is encrypted using the certificate's private key can be decrypted only with the public key. Using this scheme, information from the owner can be encrypted, digitally signed, and authenticated. IKE uses the public key to assist in verifying a host's identity.

Issuer Name
Certificate signing is a method by which a certificate is verified by a trusted third party. A signed certificate contains the subject name and key identifier of the certificate authority that signed it, known as the issuer.

Because there are many commercial and private certificate authorities, it is possible for a host to own multiple certificates that are signed by different certificate authorities. Depending on a site's policy, only certain certificate authorities might be trusted. Therefore, a z/OS IP security host might request that a peer use only certificates that are signed by a specific trusted certificate authority. Two configuration parameters refer to certificate authorities, the SupportedCertAuth parameter of the IkeConfig statement in the iked.conf file, and the CaLabel parameter of the RemoteSecurityEndpoint statement. The SupportedCertAuth parameter names a specific certificate authority that is recognized by the IKE daemon. The CaLabel parameter identifies a portion of a certificate authority hierarchy that is preferred by the local security endpoint.

For more information about the IkeConfig and RemoteSecurityEndpoint statements, see z/OS Communications Server: IP Configuration Reference.

Pre-shared key

The pre-shared key method of authentication enables a remote host to authenticate itself by providing a secret key, which is known to both hosts. This key is pre-configured by the administrator, and is used along with the Diffie-Hellman shared secret to derive cryptographic keys used to protect and authenticate data that flows during the phase 1 negotiation. The pre-shared key is a shared secret between the two IKE peers, and any host that does not know the shared key cannot enter into negotiation. IKE maintains a list of all the remote hosts that are authorized to negotiate. This list contains the identity of the remote host and the pre-shared key known to that host.

Negotiation modes for phase 1

For IKEv1, the phase 1 negotiation that takes place between two IKE peers happens in one of two modes, Main mode or Aggressive mode.

Main mode is more secure because it encrypts the identities of the two hosts that are contained in the IKE messages, but somewhat slower because more message exchanges are required. Main mode requires a total of six messages, three from the initiator and three from the responder.

Aggressive mode is faster, in that fewer messages are exchanged. Aggressive mode requires only three messages, two from the initiator and one from the responder. However, the identity of the two hosts is not protected in Aggressive mode. An IKE implementation is not required to support Aggressive mode.

The choice of mode depends in part on the security needs of the installation. If it is important that the identities of either host are not to be in clear text on the network, use Main mode. Otherwise, if both hosts support it, you can use Aggressive mode for faster, more efficient key exchange. One limitation of Aggressive mode is that the initiator can specify only a single Diffie-Hellman group because the key exchange data is sent in the first message. If you need to allow the initiator to send multiple offers with different Diffie-Hellman groups, use Main mode.

For IKEv2, there is only one set of network flows defined for phase 1 negotiation. The set requires four messages, two from the initiator and two from the responder. The first two of these four messages flow in the clear on the network; the other two messages are encrypted. The negotiation of the first phase 2 Security Association is accomplished within these four flows; when the fourth message flows, both phase 1 and phase 2 Security Associations are activated. Either the initiator or the responder can subsequently activate additional phase 2 Security Associations using this phase 1 Security Association.

You configure the choice of negotiation mode using the KeyExchangeAction statement and the Diffie-Hellman group using the KeyExchangeOffer statement in an IP security policy configuration file. For more details about the KeyExchangeAction and KeyExchangeOffer statements, see z/OS Communications Server: IP Configuration Reference.

Phase 2

The purpose of phase 2 negotiation is to establish a set of parameters that are known as a Security Association, which is used to protect specific types of IP traffic. The phase 2 Security Association contains the keys that are used to encrypt and decrypt IPSec packets on the host, authenticate IPSec packets on the host, or both. The phase 2 Security Association is negotiated for a specific set of data endpoints for a specific type of traffic, and contains the following information:

- The keys that are used to encrypt, if encryption is being used
- The keys that are used to authenticate
- The data endpoints, either a single IP address or range of IP addresses
- The protocol of the traffic to be protected, either a single protocol or all protocols
- The ports of the traffic to be protected, either a single port, a range of ports if the IKEv2 protocol is used to negotiate the phase 2 Security Association, or all ports
- The IPSec protocol that is used to protect the data: AH or ESP, or both if the IKEv1 protocol is used for the negotiation
- The type of authentication algorithm to be used
- The type of encryption algorithm to be used, if encryption is being used
- How to build the IPSec packets (tunnel, transport, UDP-encapsulated tunnel, or UDP-encapsulated transport)
- How often the keys should be refreshed, if the IKEv1 protocol is used for the negotiation
- A security parameter index (SPI) value, used together with the remote security endpoint IP address to uniquely identify the Security Association
- The Diffie-Hellman group for perfect forward secrecy (PFS)

A phase 2 negotiation can begin only after the completion of a corresponding phase 1 Security Association. The encryption methods that are agreed upon during the phase 1 negotiation are used to protect the data that is exchanged during the phase 2 negotiation. For instance, if the KeyExchangeRule between two security endpoints specified SHA1 and 3DES, the IKE data that is exchanged during the phase 2 negotiation is authenticated using HMAC-SHA and encrypted using 3DES encryption. Although both phase 1 and phase 2 Security Associations might use the same authentication and encryption methods, this is not required. For instance, a phase 1 Security Association can specify SHA1 authentication and 3DES encryption, while a phase 2 Security Association might use ESP with HMAC-MD5 authentication and DES encryption.

The keys that are generated during the phase 2 negotiation can be derived from the phase 1 master key to amortize the cost of the phase 1 key generation. As an alternative, you can configure the phase 2 negotiation to use perfect forward secrecy (PFS) for stronger security. Each has its advantage as follows:

- If PFS is used, phase 2 negotiation does not derive its keys from the master key, but instead generates new keying material using the Diffie-Hellman algorithm. Because it is independently derived, the resultant phase 2 key is more secure.
- If PFS is not used, phase 2 negotiation is completed much more quickly, but the resultant phase 2 key is less secure.

In either case, phase 2 does not incur the same degree of processing overhead that is involved in phase 1 negotiation with the remote IKE peer.

For IKEv2, the negotiation of the first phase 2 Security Association on a given phase 1 is a special case because it is performed during activation of the phase 1. In this case, the phase 1 Diffie-Hellman group is used for the phase 2 negotiation, regardless of what was defined on the IpDataOffer statement that corresponds to the phase 2.

When phase 2 negotiation has completed, the Security Association is available for use by the stack. Now, any traffic that is mapped to this Security Association by an IP filter rule is IPSec-protected. Because each phase 2 Security Association corresponds to a single unique phase 1 Security Association, the identity of the remote peer is implicitly authenticated when the phase 2 Security Association is used.

Policy for phase 2 Security Associations is defined by referencing an IpDynVpnAction statement on an IpFilterRule statement. For more details on referencing an IpDynVpnAction statement on an IpFilterRule statement, see z/OS Communications Server: IP Configuration Reference.

You can specify only tunnel or transport mode encapsulation on the IpDataOffer statement. The decision to use UDP-encapsulated transport or tunnel mode is made heuristically by the z/OS IKE daemon. If a NAT device is detected in the process of creating a phase 1 Security Association, all phase 2 Security Associations negotiated under the protection of that phase 1 Security Association are negotiated using UDP encapsulation. In this case, all IpDataOffer statements that contain ESP refer to UDP-encapsulated tunnel or transport mode, as applicable. Any IpDataOffer statements that are configured to use AH authentication are ignored, because the IPSec protocols do not allow for AH authentication when UDP encapsulation is used.

The phase 2 parameter values supported can differ between IPSec implementations. For example, z/OS provides configuration to negotiate a phase 2 Security Association for specific port and specific protocol values. Many implementations of IPSec support negotiating only a wide Security Association, covering all ports and protocols.

Refreshing phase 1 Security Associations

Refreshing a Security Association is the process of creating a new Security Association to replace an existing Security Association. The IKED automatically refreshes Security Associations when they are about to expire.

When an IKEv1 phase 1 Security Association is refreshed, the IKED performs the following actions:

- It creates a new Security Association using a main mode or aggressive mode exchange.
- It negotiates new keys and it reauthenticates the identity of the IKE peer.

When an IKEv2 phase 1 Security Association is refreshed, the IKED performs the following actions:

- It creates a new Security Association by using a create child exchange process.
- It negotiates new keys but does not reauthenticate the identity of the IKE peer.

You can use the ReauthInterval parameter on the KeyExchangeAction statement to cause the IKED to periodically reauthenticate an existing IKEv2 phase 1 Security Association. For more information about the KeyExchangeAction statement, see the KeyExchangeAction statement in z/OS Communications Server: IP Configuration Reference.

You can use the refresh option on the **ipsec** command to refresh an existing phase 1 Security Association. When you use the **ipsec** command to refresh an existing IKEv1 or IKEv2 phase 1 Security Association, new keys are negotiated and the identity of the IKE peer is reauthenticated. For more information about the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

IPSec and network address translation devices

There are inherent incompatibilities between IPSec and network address translation (NAT) functions. RFC 3715, *IPsec-Network Address Translation (NAT) Compatibility Requirements*, describes the problems that arise when IPSec is used to protect traffic that traverses a NAT. One basic problem is that when IPSec Security Associations traverse a NAT, the NAT is unable to update IP addresses and checksums that are part of the encapsulated data (encrypted, authenticated, or both). See "NATT support level" on page 933 for descriptions of the RFCs that define mechanisms to enable specific uses of IPSec to traverse one or more NAT devices.

When IPSec traverses one or more NAT devices, it is known as NAT traversal (NATT). z/OS Communications Server supports NAT traversal for IPv4 traffic only. NAT traversal is supported for IKEv1 and IKEv2 dynamic tunnels; it is not supported for manual tunnels.

There are several reasons why network address translation might be used. One reason is to economize on the use of public addresses within the internal network, using a public address only when data must be globally routed. A second reason is to hide internal IP addresses from network segments outside the internal IP address domain.

It should be noted that the NAT traversal support, as defined by the IETF, transmits internal IP addresses to its IPSec peer. These internal addresses are not exposed on the external network. However, the internal addresses are available for display at the remote security endpoint. If you are considering using this support, evaluate whether transmitting internal IP addresses to an IPSec peer is acceptable from a security policy perspective.

NAT encompasses the following IP address mapping techniques:

- One technique performs a one-to-one address translation. An internal-external IP address mapping is maintained by the NAT device. IP addresses are translated, but ports are unchanged. The mapping can be static or dynamic. For a static mapping, there is a definition in the NAT that always translates IP address *x.x.x.x* to IP address *y.y.y.y*. An outbound packet is not needed to establish the mapping. For a dynamic mapping, the NAT has a pool of IP addresses that are assigned as needed, so IP address *x.x.x.x* might be mapped to IP address *y.y.y.y* one time, and to IP address *z.z.z.z* at another time. The mapping is established when an outbound packet is processed.
- The network address port translation (NAPT) technique uses multiple internal IP addresses that are translated into a single public IP address. As part of this translation process, the TCP and UDP ports in the packets are translated. NAPT is sometimes referred to as port address translation (PAT) or IP masquerade. The mapping is typically dynamic and established in the NAPT when an outbound packet is processed. For example, the NAPT might create a mapping for internal address *x.x.x.x* port *y* to external address *a.a.a.a* port *b* and a mapping for internal address *z.z.z.z* port *v* to external address *a.a.a.a* port *c*.

  When only one client behind a NAPT has negotiated a Security Association, it is not always possible for the remote peer to detect whether the public address is being used for one-to-one address translation or for NAPT. When multiple clients have active Security Associations, the remote peer can detect when port translation is being performed.

The terms *in front of* and *behind* are used to convey which IP address a NAT is translating. When a NAT is said to be in front of a client, it means the client's address will be translated by the NAT. When a server is said to be behind a NAT, it means that server's address will be translated by the NAT.

NATT support level

z/OS Communications Server supports NAT traversal as defined in RFCs 3947, 3948, and 5996; they define mechanisms that enable IPSec to traverse one or more NAT devices. Platforms that have implemented their NAT traversal support using pre-RFC drafts might not interoperate with implementations that are compliant with RFCs 3947, 3948, and 5996.

- RFC 3947, *Negotiation of NAT-Traversal in the IKE*, allows an IKEv1 daemon to detect when one or more NATs are being traversed.
- RFC 3948, *UDP Encapsulation of IPsec ESP Packets*, defines two IPSec encapsulation modes, UDP-encapsulated tunnel mode and UDP-encapsulated transport mode. These modes facilitate the traversal of IPSec traffic through a NAT by encapsulating ESP packets within a UDP packet.
- RFC 5996, *Internet Key Exchange (IKEv2) Protocol*, specifies how to detect when IKEv2 peers are traversing one or more NATs

z/OS Communications Server does provide limited support for the following pre-RFC implementations:

- draft-ietf-ipsec-nat-t-ike-02 (pre-RFC draft of RFC 3947), and draft-ietf-ipsec-udp-encaps-02 (pre-RFC draft of RFC 3948)
- draft-ietf-ipsec-nat-t-ike-03 (pre-RFC draft of RFC 3947), and draft-ietf-ipsec-udp-encaps-03 (pre-RFC draft of RFC 3948)

Dynamic structures used to map Security Associations

The negotiation of a phase 2 Security Association results in the dynamic creation of new filters used to map the Security Association to a specific type of traffic. These dynamic filters are then added to the filter table and remain as long as the Security Association is available for use. Additional structures are added to the filter table when the remote security endpoint is behind a NAT.

Anchor filters and dynamic filters

Filters with ipsec actions are flagged as anchor filters. Anchor filters are neither permit or deny rules, but rather serve as place holders for dynamic filters in the ordered list of filter rules. A dynamic filter is an extension of an anchor filter and is created when a phase 2 Security Association is created. Each phase 2 Security Association that is negotiated is associated with two dynamic filters, an inbound filter and an outbound filter. When an IP packet matches an anchor filter rule, there is a secondary search for a matching dynamic filter rule. If one is not found, the packet is denied, unless the packet is an outbound packet and on-demand negotiations are allowed. In that case, an IKE negotiation ensues to create a Security Association and the matching dynamic filter. If a dynamic filter already exists, the action taken is to permit with IPSec processing applied. The dynamic filter rule indicates which Security Association should be used when applying IPSec processing, because there is a one-to-one correspondence between dynamic filter pairs (inbound and outbound) and phase 2 Security Associations. For a sample display of anchor and dynamic filters, see “Displaying active filters with the ipsec command” on page 1051.

NATT anchor and NATT dynamic filters

When the remote peer is behind a NAT, the dynamic anchor still serves as a place holder in the ordered list of filter rules. However, the dynamic filters that are created when a phase 2 Security Association is created are handled differently. When the phase 2 negotiation is successful, the dynamic filter pair that is created contains the 5-tuple information for which the Security Association was negotiated:

- The data endpoints, either a single IP address or range of IP addresses
- The protocol of the traffic to be protected, either a single protocol or all protocols
- The ports of the traffic to be protected, either a single port or all ports

In cases when the remote peer is behind a NAT, this 5-tuple might not be unique for a Security Association. An additional structure, a NATT anchor, is generated to anchor dynamic filters that share the same 5-tuple information. The dynamic filter is then an extension to the NATT anchor and is flagged as a NATT dynamic. For a sample display of NATT anchor and NATT dynamic filters, see "Displaying active filters with the ipsec command" on page 1051.

NAT resolution filters

When the remote peer is behind a NAT, a connection level filter structure is also created for TCP and UDP connections, the NAT resolution filter (NRF). Unlike the NATT anchor and NATT dynamic filters, which are created when the phase 2 Security Association is created, the NRF is created when the first inbound packet is received for the connection over the phase 2 Security Association. The NRF contains a single source and destination IP address, a single source and destination port value, and a single protocol. This connection-level filter is needed to determine the phase 2 Security Association that will be used for outbound data. For a sample display of NAT resolution filters, see "Displaying active filters with the ipsec command" on page 1051.

Remote port translation

When the remote peer is a security gateway behind a NAT, the remote data endpoint of the client is represented by the security gateway's public IP address. From the z/OS server's perspective, multiple clients that are behind the security gateway are all identified by the IP address of their corresponding security gateway, not the client's private IP address. Consequently, if two clients behind the security gateway were to open a connection to the same service (FTP, for example) from the same ephemeral port (1024, for example), the two connections could not be uniquely identified.

Similarly, when the remote peer is a host behind a NAPT, the remote data endpoint of the host is represented by the public address assigned by the NAPT. Typically, multiple hosts that are behind the NAPT are all assigned the same public IP address, with port translation distinguishing the hosts. The port that is being translated by the NAPT is the port found in the UDP header of the IKE packet and the UDP-encapsulated ESP packet, not the connection port. Consequently, if two hosts behind the NAPT open a connection to the same service (FTP, for example) from the same ephemeral connection port (1024, for example), the two connections cannot be uniquely identified.

Traditionally, the combination of local and remote IP addresses and ports is enough to distinguish the connections; in these instances, because the connection requests appear to be coming from the same IP address, they look identical. To avoid such conflicts, Communications Server provides remote port translation for TCP and UDP connections when needed. Without remote port translation, the first inbound TCP connection request for a unique 5-tuple would succeed. However, a subsequent inbound connection request using the same 5-tuple would fail. Communications Server's remote port translation allows subsequent inbound connections using the same 5-tuple to succeed by translating the remote port value to a different ephemeral port.

Remote port translation does not need to be configured or enabled. It is built into Communications Server's NAT traversal support and is always in effect when the remote peer is behind a NAT. When a remote port is translated, message EZD0827I, remote port translated, is logged. Remote port translation determines whether the source port in an inbound packet represents a duplicate port. If it does, an attempt is made to assign an unused ephemeral port value. If the inbound packet matches a configured filter rule that covers all ports, any unused ephemeral port value can be assigned. If the configured filter rule applies only to a range of ports, the remote port is translated only to an unused ephemeral port in that range. If an unused port cannot be found, the connection request fails. When a remote port value is translated, there is both an original remote port and translated remote port for a connection. Various commands, such as Netstat, enable you to view connection information including the port, or even to select display data based on port. For connection data, if only one remote port is being provided, the translated port is displayed or used for a selection, if remote port translation has been done.

The **ipsec -o display** command provides a display of port mappings in effect. For a sample display of remote port translation, see "Displaying remote port translation with the ipsec command" on page 1065.

Steps for preparing the z/OS system for IP security

Establish the security needs for your installation. Set up key required applications, modify default IP filter policy if necessary, configure the IKE daemon, configure the NSS daemon and additional encryption products as needed, and create IP security policy configuration files.

Before you begin

You need to be familiar with the concepts of IP filter logging and IPSec protection. See "Overview of using IP security" on page 888. You also need to be familiar with the commands used to administer IP security. See "Commands used to administer IP security" on page 887.

Procedure

Perform the following steps to prepare the z/OS system for IP security.

1. Develop a site policy to address the security needs of your installation.

   Consider the following questions before beginning to build a robust IP security policy:

   - Do you want a default-allow or default-deny policy?
   - If you want a default-deny policy, what traffic is allowable as critical to the correct functioning of the system?
   - What hosts are allowed to send data to the secure host?
   - Do you understand the network topology, and where the communicating hosts are located within the network?
   - Are there network address translation (NAT) devices in the network topology?
   - What type of traffic is allowable from those hosts?
   - What general network services that rely on well-known ports do you want to allow?
   - Will you forward any packets you receive that are not destined for you?
   - Who is allowed to configure the IP security policy?
   - Are you running with one TCP/IP stack, or more than one TCP/IP stack?
   - Are you running IPv6 traffic?
   - Will all TCP/IP stacks share the same policy definitions, or will they have unique differences?
   - Will traffic that is sent or received by the secure host traverse the Internet?
   - Do you want to participate in a VPN? (If so, IPSec is required.)
   - If IPSec is required:
     - What are the IPSec capabilities of the remote endpoints?
     - How do you want to authenticate participating hosts:
       - Pre-shared key?
       - Digital signature? (requires digital certificates)
     - What type of authentication is needed for the data that is involved?
     - Is encryption needed for the data that is involved, and how strong should the encryption algorithms be?
     - Will all participating hosts use the same level of data encryption and authentication, or do you need to define unique policies for individual hosts?
2. Identify the resources to be secured.

   Create a worksheet for each TCP/IP stack you will enable for IP security. This information aids in determining which interfaces and connected networks are secured. You can later use the information from the worksheets to provide values for IP security policy configuration statements. Figure 122 on page 936 includes a sample worksheet:

```
Host name of z/OS system _____________________________________

Complete for each TCP/IP stack on this host:
TCP/IP stack name _____________IPSECURITY-enabled (Y/N)______
Network interface(s)
IPv4 address/Mask_____________________Security Class_________
IPv4 address/Mask_____________________Security Class_________
IPv4 address/Mask_____________________Security Class_________
IPv4 address/Mask_____________________Security Class_________
IPv6 address/Prefix___________________Security Class_________
IPv6 address/Prefix___________________Security Class_________
Virtual IPv4 address/Mask_____________________
Virtual IPv4 address/Mask_____________________
Virtual IPv4 address/Mask_____________________
Virtual IPv4 address/Mask_____________________
Virtual IPv6 address/Prefix___________________
Virtual IPv6 address/Prefix___________________

Identities, other than IP address, by which the IKE daemon will be known
(e.g., X500dn, Fqdn, UserAtFqdn, KeyID)
___________________
___________________
___________________
___________________
___________________
```

Figure 122. Sample worksheet for stack security

Security class is an optional designation for a network interface. You can assign network interfaces a security class (1-255) that is used to group interfaces with similar security requirements. This concept is an extension of the traditional notion of secure and nonsecure interfaces, to allow for more than two classes. The secure and nonsecure model can still apply if only two security classes are defined.

Complete a table of remote hosts and subnetworks with which this host needs to transfer data, as shown in Table 50 on page 936. The remote hosts can optionally be grouped in a user-defined zone to simplify the number of IP filter rules that are required. For instance, you might want to define an internal, external, trusted, Internet, partner company, or other zone that has meaning to your site.

Rule: If the remote IKE peer is on a security gateway to the remote host, the IP address of the remote host might not match the IP address of the remote IKE peer. If this is the case, tunnel-mode encapsulation of IPSec traffic is required.

Table 50. Table of remote hosts and subnetworks

Remote IP address or subnet	User-defined zone of remote hosts	Identity of remote IKE peer (IpAddr, X500dn, KeyID, Fqdn, or UserAtFqdn)	IP address of remote IKE peer	Allowed list of services (Telnet, FTP, web, EE, ALL, or other)	Security action (permit, deny, ipsec)	IPSec security level, if applicable

3. Modify the default IP filter policy (optional).

   After the site policy is established, you can modify the default IP filter policy if necessary. For a description and examples of how to modify the default IP filter policy, see “Default IP filter policy and IP security policy” on page 901.

   Guideline: You should define IPSECRULE and IPSEC6RULE statements that permit access from at least one administrative machine. In the event that Policy Agent is unavailable, or the IP security configuration files contain errors, the active default policy would deny access to the IP security-enabled stack. Should this situation occur, if you have added the appropriate IPSECRULE and IPSEC6RULE statements to the default policy, you can still access the secure z/OS host and make the necessary administrative changes to correct the problem.

4. Set up the key required applications:

- TCP/IP

  To enable IP security on a z/OS stack, make the following changes in the TCP/IP profile:

  – Add IPCONFIG IPSECURITY.
  – To also enable IP security for IPv6, add IPCONFIG6 IPSECURITY.
  – Reserve ports 500 and 4500 for IP security. If the IKE daemon is running on this system, reserve the ports for the user ID under which the IKE daemon is running. In this example, the IKE daemon is running under the IKED user ID:

    ```
    500   UDP        IKED
    4500  UDP        IKED
    ```

    If the IKE daemon is not running on this system, reserve the ports by specifying RESERVED:

    ```
    500   UDP        RESERVED
    4500  UDP        RESERVED
    ```

  – To direct IPSec's AH and ESP protocol processing to zIIPs, add GLOBALCONFIG ZIIP IPSECURITY.

  For information on the IPCONFIG, IPCONFIG6, GLOBALCONFIG, and PORT statements, see z/OS Communications Server: IP Configuration Reference.

- Policy Agent

  For information on configuring the Policy Agent, see Chapter 14, “Policy-based networking,” on page 787.

- TRMD

  For information on configuring TRMD, see “TRMD” on page 879.

- Syslogd

  For information on configuring the syslog daemon, see “Configuring the syslog daemon” on page 213.

- IKED

  The IKED can be started from a z/OS UNIX command line or as an MVS procedure. The iked.conf file controls the overall function of the IKE daemon, such as the following settings:

  – The logging level of the IKE daemon
  – The logging level of the Policy Agent when performing IP security-related tasks on behalf of the IKE daemon
  – The name of the RACF key ring that is owned by the IKE daemon
  – Whether IKE messages are echoed to STDOUT when the daemon is started for the UNIX System services shell
  – How long to wait when attempting to connect to the Policy Agent
  – The list of certificate authorities that are acceptable for RSA signature negotiation when the IKED is using the native certificate service (when the Network Security Services (NSS) server is providing the certificate service, this list is not used)

  Most of the configuration parameters have a default value and do not require modification.

  Rule: If the RSA signature method is to be used in any IKE phase 1 negotiation and the IKED is using the native certificate service, the name of the IKE key ring must be included in the iked.conf file. If the NSS server is providing the certificate service, the IKE key ring name is not needed.

  Tip: If the RSA signature method is used and the IKED is using the native certificate service, including a list of supported certificate authorities enhances the performance of certificate searches.

  Guideline: You should not change the logging levels of the IKE daemon (IkeSyslogLevel) and Policy Agent (PagentSyslogLevel) from their default values for normal day-to-day operation. Higher logging levels can affect performance and should be used for temporary diagnostic purposes

only. The default PagentSyslogLevel value 0 prevents the IKE daemon from logging diagnostic information about its interactions with Policy Agent. The default IkeSyslogLevel value 1 provides basic informational and error messages. You can set the IkeSyslogLevel value to 0 to disable IKE syslog messages entirely (for both the IKE daemon and Policy Agent; to collect PagentSyslogLevel tracing, the IkeSyslogLevel value must also be a nonzero value). You can set the IkeSyslogLevel value to a higher value to identify the source of an error; for example, if you experience problems with Security Association negotiations that are caused by a configuration error, you might enable IkeSyslogLevel 4 (debugging information for Security Association negotiations).

For detailed syntax and a description of the iked.conf file, and details on starting the IKE daemon as an MVS procedure, see z/OS Communications Server: IP Configuration Reference.

5. Define access controls for key required applications:
   - TRMD

     TRMD runs as an authorized program and requires RACF setup. TRMD must be able to run as a started task and have superuser authority. For sample RACF commands, see the EZARACF member of SEZAINST.
   - Policy Agent

     Policy Agent runs as an authorized program and requires RACF setup. Policy Agent must be able to run as a started task and have superuser authority. For sample RACF commands, see the EZARACF member of SEZAINST and "Step 1: Configure general information" on page 808.
   - IKED

     For the steps to prepare for running the IKE daemon, see Appendix E, "Steps for preparing to run IP security," on page 1361.
6. Configure the IKE daemon.

   See Appendix E, "Steps for preparing to run IP security," on page 1361.
7. (Optional) Configure the NSS daemon.

   If the IKED is using the NSS certificate service for any IKEv1 or IKEv2 Security Association negotiation, then configure the NSS daemon. The IKED must use the NSS certificate service for any IKEv2 negotiation that uses certificates for authentication. The IKED can use either the native certificate service or the NSS certificate service for an IKEv1 negotiation that uses certificates for authentication.

   For more information about the NSS daemon, see Chapter 18, "Network security services," on page 1083.
8. (Optional) Configure additional encryption products.
   - 3DES support

     For 3DES (triple DES), the Communications Server Security Level 3 feature, is required.
   - AES support (including AES-CBC, AES-GCM and AES-GMAC)

     For AES, the Communications Server Security Level 3 feature and the z/OS Security Level 3 feature are required and ICSF must be started.
   - FIPS 140

     For IKED, NSSD, and TCP/IP to run in FIPS 140 mode, you must first choose one of several modes of FIPS operation for ICSF and start ICSF.

     Requirement: ICSF must be active before starting the IKE daemon or NSS server configured in FIPS 140 mode. For information about enabling ICSF to support FIPS 140-2, see the topic about operating in compliance with FIPS 140-2 in z/OS Cryptographic Services ICSF Writing PKCS #11 Applications.
   - ICSF

     There are several options available on z/OS to perform encryption in hardware. The ICSF product is required to support these various options. For a description and information about how to configure these options, see z/OS Cryptographic Services ICSF Administrator's Guide.

For the RACF commands that authorize ICSF, see Appendix E, “Steps for preparing to run IP security,” on page 1361.

- SHA2 support (including SHA2-256, SHA2-384, and SHA2-512, as well as all corresponding HMAC-SHA2 algorithms) and AES authentication (AES-XCBC)

  For SHA2 and AES-XCBC authentication, ICSF must be started.

9. Create IP security policy configuration files.

   After the security needs of your installation are established, the next step is to create one or more IP security policy configuration files. There are two main configuration files in which IP security policy configuration is stored:

   Stack-specific IP security configuration file
   Configured on a per-stack basis and contains only IP security policy configuration that applies to the stacks for which it is configured.

   Common IP security configuration file
   Contains IP security policy configuration that applies to every stack on the system and can be used to hold shared definitions.

   The stack-specific file can refer to various definitions in the common file, and can override policy definitions in the common file.

   Although Policy Agent uses LDAP to store policy for some other policy types, IP security policy configuration is stored exclusively in human-readable text files, either in the z/OS UNIX file system or in an MVS data set. These configuration files are then read by Policy Agent when it initializes and before being installed into the stack.

IP security policy configuration

This topic describes the general steps for creating IP security policy configuration files for the most common configurations. Configuring a complete and specific IP security policy that meets the needs of any installation is beyond the scope of this text, but guidance for more advanced configurations is provided.

“Overview of configuring IP security policy” on page 939 describes the common IP security configuration file, the stack-specific IP security configuration file, and the general content, structure, and use of these files.

“Component policies of IP security policy configuration files” on page 946 describes the types of policies contained in IP security policy configuration files.

“Steps for configuring IP security policy” on page 972 describes the steps for manually creating IP security policy configuration files.

“Quick start using IP filtering and IPSec host-to-host” on page 957 describes a complete IP security policy allowing connections from a secure server to an administrative machine on an internal network, and represents the minimum configuration needed to provide IPSec protection with dynamic key management between two hosts. This topic also describes the use of the **ipsec** command to display filters and Security Associations.

“Configuring specific security models” on page 974 provides more examples and describes the configuration needed for common security models.

Overview of configuring IP security policy

There are three options for IP security policy configuration for a system:

- Use a common IP security configuration file that applies to all stacks on the system, enforcing a consistent policy. In this instance, a stack-specific IP security configuration file is not necessary.
- Use a unique and separate stack-specific IP security configuration file for each stack on the system. In this instance, a common IP security configuration file is not necessary.
- Use both a common and a stack-specific IP security configuration file.

- The common IP security configuration file can be used as a common repository for frequently used definitions, which can be referenced by any stack-specific IP security configuration file.
- The stack-specific IP security configuration file can contain unique statements that apply only to the stack for which it is configured, and can reference statements that are defined in the common IP security configuration file.

Although not an error, note that when using the last approach, it is possible for duplicate statements to exist in the common and the stack-specific IP security configuration files (for example, two IpFilterRule statements with the same name). In this case, the statement in the stack-specific IP security configuration file is honored. Statements in the stack-specific IP security configuration file always take precedence over the common IP security configuration file.

Structure of an IP security configuration file

The common IP security configuration file and the stack-specific IP security configuration file have exactly the same structure. They are comprised of a number of statements that define items that are used to define policy, such as policies, rules, actions, groups, and objects. Statement names and attribute names are not case sensitive, though they appear in mixed case in this information for readability. Only user-defined names are case sensitive. For the complete syntax of all IP security policy statements, see z/OS Communications Server: IP Configuration Reference.

An IP security policy configuration statement has the following generic form:

```
StatementType     user-defined name
{
   Attribute1     value1
   Attribute2     value2
   .
   .
   .
}
```

Statements often contain other inline statements in a recursive form:

```
StatementType1     user-defined name
{
   Attribute1      value1
   StatementType2  optional user-defined name
   {
      Attribute1   value1
      Attribute2   value2
   }
   Attribute2      value2
}
```

There are three main sections in an IP security configuration file, identified by the following three statements:

- IpFilterPolicy
- KeyExchangePolicy
- LocalDynVpnPolicy

Additional statements that define rules, actions, groups, and objects are found both in the main body of the configuration file and within any of these other three policy blocks. A high-level view of an IP security configuration file follows. Although the statement blocks are shown in a specific order, the ordering is arbitrary.

```
IpFilterPolicy #(required)
{
  <local statements>
}

KeyExchangePolicy #(optional)
{
  <local statements>
}
```

```
LocalDynVpnPolicy #(optional)
{
  <local statements>
}

  <global statements>
```

Groups

Groups provide a method to combine related objects in a meaningful way into sets. The following IP security policy configuration statements can be used as groups:

- IpAddrGroup
- IpFilterGroup
- IpServiceGroup
- KeyExchangeGroup
- LocalDynVpnGroup

Reference statements

A reference statement can be recognized by the suffix Ref. References provide a convenient way to reuse definitions, eliminating the need to repeatedly specify things such as host addresses and common services. Nearly all statements in the IP security configuration file can be referenced, and all action statements must be referenced. To be referenced, an IP security policy configuration statement must be given a user-defined name. Names of statements can be up to 32 characters. The following IP security policy configuration statements can be referenced, and therefore reused:

- IpAddr
- IpAddrSet
- IpDataOffer
- IpDynVpnAction
- IpFilterGroup
- IpFilterRule
- IpGenericFilterAction
- IpLocalStartAction
- IpManVpnAction
- IpService
- IpServiceGroup
- IpTimeCondition
- KeyExchangeAction
- KeyExchangeGroup
- KeyExchangeOffer
- KeyExchangeRule
- LocalDynVpnGroup
- LocalDynVpnRule
- LocalSecurityEndpoint
- RemoteIdentity
- RemoteSecurityEndpoint

Steps for configuring local IP security policy using only a common IP security configuration file

You can configure local IP security policy for all stacks on the system using only a common IP security configuration file. In this instance, a stack-specific IP security configuration file is not necessary.

Procedure

Perform the following steps to configure local IP security policy using only a common IP security configuration file.

1. In the main Policy Agent configuration file, include a CommonIpSecConfig line that identifies the common IP security configuration file, as follows:

```
CommonIpSecConfig      /etc/common.ipsecpol
```

2. In the main Policy Agent configuration file, include a line with the TcpImage statement for each IP security stack to be configured:

```
TcpImage TCPCS  /etc/TCPCS.image
TcpImage TCPCS2 /etc/TCPCS2.image
⋮
```

3. In each configuration file that was identified on the TcpImage statement shown in step “2” on page 942, include a line that contains IpSecConfig with no file name, as follows:

```
In /etc/TCPCS.image:
IpSecConfig

In /etc/TCPCS2.image:
IpSecConfig
```

Results

All stacks on the z/OS system will adhere to the policy that is specified in the /etc/common.ipsecpol file.

Steps for configuring remote IP security policy using only a common IP security configuration file

You can configure remote IP security policy for all stacks on the z/OS policy client system using only a common IP security configuration file. In this instance, a stack-specific IP security configuration file is not necessary.

Procedure

Perform the following steps to configure remote IP security policy using only a common IP security configuration file.

1. In the main Policy Agent configuration file on the policy client, include the ServerConnection statement, and a line with the TcpImage statement for each IP security stack to be configured:

```
ServerConnection
{
    …
}
TcpImage TCPCS  /etc/TCPCS.image
TcpImage TCPCS2 /etc/TCPCS2.image
⋮
```

2. In each configuration file that was identified on the TcpImage statement shown in step 1, include a PolicyServer statement.

   For example, in /etc/TCPCS.image:

```
PolicyServer
{
   ClientName  IPSecClientTCPCS
   PolicyType IPSec
   {
      …
   }
   …
}
```

In /etc/TCPCS2.image:

```
PolicyServer
{
   ClientName  IPSecClientTCPCS2
   PolicyType IPSec
   {
      …
   }
   …
}
```

3. In the main configuration file on the policy server, include a DynamicConfigPolicyLoad statement, as follows:

```
DynamicConfigPolicyLoad  IPSecClient.*
{
   PolicyType IPSec
   {
      CommonPolicyLoad /etc/common.ipsecpol
   }
   …
}
```

Results

All stacks on the z/OS policy client system will adhere to the policy that is specified in the /etc/common.ipsecpol file on the policy server.

Steps for configuring local IP security policy using only a stack-specific IP security configuration file

You can configure local IP security policy using only a unique and separate stack-specific IP security configuration file for each stack on the system. In this instance, a common IP security configuration file is not necessary.

Procedure

Perform the following steps to configure local IP security policy using only a stack-specific IP security configuration file.

1. In the main Policy Agent configuration file, include a line with the TcpImage statement for each stack to be configured, as follows:

```
TcpImage TCPCS  /etc/TCPCS.image
TcpImage TCPCS2 /etc/TCPCS2.image
⋮
```

2. In each configuration file that was identified on the TcpImage statement shown in step “1” on page 943, include an IPSecConfig line that identifies the stack-specific IP security configuration file, as follows:

```
In /etc/TCPCS.image:
IpSecConfig     /etc/TCPCS.ipsecpol

In /etc/TCPCS2.image:
IpSecConfig     /etc/TCPCS2.ipsecpol
```

Results

Each stack on the z/OS system will adhere to the policy that is specified by its unique policy file. Stack TCPCS uses the policy that is configured in /etc/TCPCS.ipsecpol, and stack TCPCS2 uses the policy that is configured in /etc/TCPCS2.ipsecpol.

Steps for configuring remote IP security policy using only a stack-specific IP security configuration file

You can configure remote IP security policy using a unique and separate stack-specific IP security configuration file for each stack on the z/OS policy client system. In this instance, a common IP security configuration file is not necessary.

Procedure

Perform the following steps to configure remote IP security policy using only a stack-specific IP security configuration file.

1. In the main Policy Agent configuration file on the policy client, include the ServerConnection statement, and a line with the TcpImage statement for each IP security stack to be configured:

```
ServerConnection
{
   …
}
TcpImage TCPCS  /etc/TCPCS.image
TcpImage TCPCS2 /etc/TCPCS2.image
⋮
```

2. In each configuration file that was identified on the TcpImage statement shown in step 1, include a PolicyServer statement.

   For example, in /etc/TCPCS.image:

```
PolicyServer
{
   ClientName  IPSecClientTCPCS
   PolicyType IPSec
   {
      …
   }
   …
}
```

   In /etc/TCPCS2.image:

```
PolicyServer
{
   ClientName  IPSecClientTCPCS2
   PolicyType IPSec
   {
      …
   }
   …
}
```

3. In the main configuration file on the policy server, include DynamicConfigPolicyLoad statements, as follows:

```
DynamicConfigPolicyLoad  IPSecClientTCPCS
{
   PolicyType IPSec
   {
      PolicyLoad /etc/TCPCS.ipsecpol
   }
   …
}
DynamicConfigPolicyLoad  IPSecClientTCPCS2
{
   PolicyType IPSec
   {
```

```
      PolicyLoad /etc/TCPCS2.ipsecpol
   }
   …
}
```

Results

Each stack on the z/OS policy client system will adhere to the policy that is specified by its unique policy file. Stack TCPCS uses the policy that is configured in /etc/TCPCS.ipsecpol on the policy server, and stack TCPCS2 uses the policy that is configured in /etc/TCPCS2.ipsecpol on the policy server.

Steps for configuring local IP security policy using both a stack-specific file and a common file

You can configure local IP security policy using both a stack-specific file and a common file. If there are duplicate statements, the stack-specific file always takes precedence over the common file.

Procedure

Perform the following steps to configure local IP security policy using both a stack-specific IP security configuration file and a common IP security configuration file.

1. In the main Policy Agent configuration file, include a CommonIpSecConfig line that identifies the common IP security configuration file, as follows:

   ```
   CommonIpSecConfig    /etc/common.ipsecpol
   ```

2. In the main Policy Agent configuration file, include a line with the TcpImage statement for each stack to be configured, as follows:

   ```
   TcpImage TCPCS  /etc/TCPCS.image
   TcpImage TCPCS2 /etc/TCPCS2.image
   ⋮
   ```

3. In each configuration file that was identified on the TcpImage statement in step “2” on page 945, include an IPSecConfig line that identifies the stack-specific IP security configuration file, as follows:

   ```
   In /etc/TCPCS.image:
   IpSecConfig     /etc/TCPCS.ipsecpol

   In /etc/TCPCS2.image:
   IpSecConfig     /etc/TCPCS2.ipsecpol
   ```

Results

Any statements in the common IP security configuration file are added to the policy for each stack when the policy is initialized. Either file, /etc/TCPCS.ipsecpol or /etc/TCPCS2.ipsecpol, can refer to statements in /etc/common.ipsecpol. In the case of duplicate names, any named statement in the stack-specific IP security configuration file overrides a statement with the same name in the common IP security configuration file.

Steps for configuring remote IP security policy using both a stack-specific file and a common file

You can configure remote IP security policy using both a stack-specific file and a common file. If there are duplicate statements, the stack-specific file always takes precedence over the common file.

Procedure

Perform the following steps to configure remote IP security policy using both a stack-specific IP security configuration file and a common IP security configuration file.

1. In the main Policy Agent configuration file on the policy client, include the ServerConnection statement, and a line with the TcpImage statement for each IP security stack to be configured:

```
ServerConnection
{
   …
}
TcpImage TCPCS  /etc/TCPCS.image
TcpImage TCPCS2 /etc/TCPCS2.image
⋮
```

2. In each configuration file that was identified on the TcpImage statement shown in step 1, include a PolicyServer statement.

   For example, in /etc/TCPCS.image:

```
PolicyServer
{
   ClientName  IPSecClientTCPCS
   PolicyType IPSec
   {
      …
   }
   …
}
```

   In /etc/TCPCS2.image:

```
PolicyServer
{
   ClientName  IPSecClientTCPCS2
   PolicyType IPSec
   {
      …
   }
   …
}
```

3. In the main configuration file on the policy server, include DynamicConfigPolicyLoad statements, as follows:

```
DynamicConfigPolicyLoad  IPSecClientTCPCS
{
   PolicyType IPSec
   {
      CommonPolicyLoad /etc/common.ipsecpol
      PolicyLoad /etc/TCPCS.ipsecpol
   }
   …
}
DynamicConfigPolicyLoad  IPSecClientTCPCS2
{
   PolicyType IPSec
   {
      CommonPolicyLoad /etc/common.ipsecpol
      PolicyLoad /etc/TCPCS2.ipsecpol
   }
   …
}
```

Results

Any statements in the common IP security configuration file are added to the policy for each stack when the policy is initialized. Either file, /etc/TCPCS.ipsecpol or /etc/TCPCS2.ipsecpol, can refer to statements in /etc/common.ipsecpol. In the case of duplicate names, any named statement in the stack-specific IP security configuration file overrides a statement with the same name in the common IP security configuration file.

Component policies of IP security policy configuration files

There are three types of policies in IP security policy configuration files:

- IP filter policy (IpFilterPolicy statement)
- Key exchange policy (KeyExchangePolicy statement)
- Local dynamic VPN policy (LocalDynVpnPolicy statement)

IP filter policy

An IP filter policy can stand alone to provide IP filtering and IPSec protection with manual key management. Used in conjunction with the two other policies, it is also required to provide IPSec protection with dynamic key management (IKE). Because filtering is crucial to secure traffic on a host, an IP security policy that contains no IpFilterPolicy statement block or an empty IpFilterPolicy statement block is considered an error, leaving the default policy that is provided by the stack in effect.

The IpFilterPolicy statement block consists of:

- A set of global configuration options
- An ordered list of IP filter rules (IpFilterRule statements)

The purpose of the global configuration options is to control global policy items, such as whether logging is active or whether on-demand Security Association negotiations are allowed, and so forth. These global options apply to all of the IP filter rules that are contained in the policy. Each IP filter rule, in turn, contains data endpoints, traffic descriptions, and actions. When a packet entering or leaving the system matches the data endpoints and traffic description in an IP filter rule, the associated action is taken. If the action is an ipsec action, additional action statements are coded that define the parameters of the IPSec Security Association.

The following sample IpFilterRule statement allows web traffic on an internal server, and references a description of each line in the sample:

```
1  IpFilterRule             InternalNetWeb
2       {
3          IpSourceAddr           9.1.1.1
4          IpDestAddrSet          9.1.1.0/24
5          IpService
6          {
7             SourcePortRange         80
8             DestinationPortRange    1024 65535
9             Protocol                tcp
10            Direction               bidirectional InboundConnect
11            Routing                 local
12            SecurityClass           0
13         }
14         IpGenericFilterActionRef   permit-nolog
15      }
```

Line
Description

1
The IpFilterRule keyword, followed by a required user-defined name for this rule.

2
An open brace ({) marks the start of an IpFilterRule statement block.

3
The source address of the rule. Outbound IP packets that match this rule must have 9.1.1.1 as the source address in the IP header.

4
The destination address of the rule. Outbound IP packets that match this rule must have an address in the range of 9.1.1.0 - 9.1.1.255 as the destination address in the IP header.

5
The IpService statement block describes the type of traffic that is allowed between the two data endpoints. The IpService block in this rule is inline to the rule, meaning that the entire definition of the IP service is included within the rule. Many policy statements, including the IpService statement, can be referenced rather than included inline.

6
An open brace ({) marks the start of the IpService statement block.

7
The range of source ports that is allowed for outbound packets. The value can be a single number, or a range of ports.

8
The range of destination ports that is allowed for outbound packets. The value can be a single number, or a range of ports.

9
The specific protocol that is allowed by this rule.

10
The direction specification for the IP packet.

Bidirectional indicates that this rule allows outbound traffic from local address 9.1.1.1 on local port 80 to any address in subnet 9.1.1.0/24 using any ephemeral port (that is, 1024-65535), and inbound traffic from any address in subnet 9.1.1.0/24 using any ephemeral port to local address 9.1.1.1 on local port 80. Without the use of the bidirectional keyword, it would be necessary to create two filter rules, one for outbound traffic and one for inbound traffic.

InboundConnect indicates that the rule will match inbound TCP connection attempts as well as bidirectional data on an established connection, but it will not match outbound TCP connection attempts.

11
The routing information for a packet matching this rule. For a packet to match this rule, the traffic must be local. In other words, this rule does not allow any packets that must be forwarded to another network node. Possible values are local, routed, or either.

12
The security class of the interface on which the packet must arrive or leave. The security class of an interface is defined using the SECCLASS parameter on the statement that is used to define the interface in the TCP/IP profile (that is, the LINK, INTERFACE, IPCONFIG DYNAMICXCF, or IPCONFIG6 DYNAMICXCF statement). Each interface on the system can be assigned a SECCLASS in the range 1 – 255. If the SECCLASS parameter is not coded in the TCP/IP profile for an interface, by default the interface is SECCLASS 255. In the SecurityClass parameter in this example, the value 0 indicates that a packet that matches this rule is not restricted and can traverse any interface (1 – 255).

13
A close brace (}) marks the end of the IpService block.

14
The action that is taken on a packet that matches this rule. In this case, permit the packet and log all occurrences of a match. All IP filter rules must include an IpGenericFilterActionRef statement. The IpGenericFilterAction statement itself should be defined elsewhere, outside of the IpFilterRule block, either in the common or the stack-specific IP security configuration file:

```
IpGenericFilterAction     permit-nolog
{
  IpFilterAction          permit
  IpFilterLogging         no
}
```

15
A close brace (}) marks the end of the IpFilterRule statement block.

Rule: Do not include policy action statements inline. They must be referenced.

Example 1

Permit rule allowing outbound FTP client connections from the local host (9.1.1.1) to a remote FTP server (9.1.1.2):

```
IpFilterRule              FTP-client
{
      IpSourceAddr           9.1.1.1
      IpDestAddr             9.1.1.2
      IpService
      {
        SourcePortRange      1024 65535
        DestinationPortRange 21
        Protocol             tcp
        Direction            bidirectional OutboundConnect
        Routing              local
        SecurityClass        0
      }
      IpService
      {
        SourcePortRange      1024 65535
        DestinationPortRange 20
        Protocol             tcp
        Direction            bidirectional InboundConnect
        Routing              local
        SecurityClass        0
      }
      IpGenericFilterActionRef  permit
}
```

Normal (non-passive mode) FTP requires that the FTP client be allowed to initiate outbound connections to port 21, and be able to receive inbound connections from port 20. The `IpGenericFilterAction permit` block must be defined elsewhere, in either the common or the stack-specific IP security configuration file:

```
IpGenericFilterAction     permit
{
  IpFilterAction          permit
}
```

Example 2

Deny rule that blocks all traffic from all private address spaces that is inbound to a public interface:

```
IpFilterRule              deny-private
{
      IpSourceAddrGroupRef   PrivateAddrs
      IpDestAddr             all
      IpService
      {
        SourcePortRange      0
        DestinationPortRange 0
        Protocol             all
        Direction            inbound
        Routing              either
        SecurityClass        0
      }
      IpGenericFilterActionRef  deny-log
}
```

The IpSourceAddrGroupRef parameter references an IP address group that is presumed to be defined elsewhere, in either the common or the stack-specific IP security configuration file:

```
IpAddrGroup PrivateAddrs
{
   IpAddrSet
   {
     Prefix 10.0.0.0/8
   }
   IpAddrSet
   {
     Prefix 172.16.0.0/12
   }
   IpAddrSet
```

```
   {
     Range 192.168.0.0-192.168.255.255
   }
}
```

The IpGenericFilterActionRef parameter references an IpGenericFilterAction statement that is presumed to be defined elsewhere, in either the common or the stack-specific IP security configuration file:

```
IpGenericFilterAction     deny-log
{
  IpFilterAction          deny
  IpFilterLogging         yes
}
```

Example 3

An ipsec rule that requires IPSec protection for all traffic between the secure server and an administrative machine on the internal network:

```
IpFilterRule                 Rule2Admin
{
     IpSourceAddrRef            InternalServerAddressA1
     IpDestAddrRef              AdminClient
     IpServiceRef               All-traffic-local
     IpGenericFilterActionRef   ipsec
     IpDynVpnActionRef          Silver-TransportMode
}
```

The use of multiple references in this example makes the IP filter rule easier to read. For each referenced object or action, there should be a corresponding definition elsewhere, in either the common or the stack-specific IP security configuration file:

```
IpAddr                 InternalServerAddressA1
{
  Addr                 9.1.1.1
}

IpAddr                 AdminClient
{
  Addr                 9.1.1.2
}

IpService              All-traffic-local
{
  Protocol             all
  Direction            bidirectional
  Routing              local
  SecurityClass        0
}

IpGenericFilterAction  ipsec
{
  IpFilterAction       ipsec
  IpFilterLogging      yes LogDeny
}

IpDynVpnAction         Silver-TransportMode
{
  Initiation           either
  InitiateWithPfs      None
  AcceptablePfs        None
  IpDataOfferRef       SHA-DES-Transport
}

IpDataOffer            SHA-DES-Transport
{
  HowToEncap           transport
  HowToEncrypt         DES
  HowToAuth            ESP HMAC_SHA1
}
```

IP filter rule order

"Example 1" on page 949, "Example 2" on page 949, and "Example 3" on page 950 show individual IP filter rules. A complete IP filter policy contains any number of IP filter rules, configured in much the same manner. It is important to remember that IP filter rules in an IP filter policy are searched in the order listed. Because it is possible for a packet to match more than one rule, a search for a matching filter rule stops after the first match is found, even if there are additional matches further down in the list. Use the **ipsec** command traffic test option (**ipsec -t**) as an aid in determining which IP filter rule an IP packet matches.

The command-line arguments to the **ipsec -t** command are a set of characteristics that describe a particular IP packet. The existing set of filter rules are searched for potential matches. Unlike normal filter processing, which stops the search after a match is found, the **ipsec -t** command displays all matching filter rules. Input to the **ipsec -t** command does not have to specify all possible filtering criteria from an IP packet. The output of the **ipsec -t** command must be inspected to determine which of the returned rules match for a given case.

For instance, an IP filter rule for ICMP can be configured for a specific type and code value, while the traffic test does not provide ICMP type and code as inputs. If more than one IP filter rule matches on the ICMP protocol, they are all displayed. You must determine, from among those listed, which rule applies for a specific IP packet.

For a complete description of the **ipsec** command, including the **ipsec -t** option, see z/OS Communications Server: IP System Administrator's Commands.

Key exchange policy

A key exchange policy is required by IKE to provide dynamic key management. The policy contains the definitions about how the negotiation of keys is to be performed (using IKEv1 or IKEv2), how the negotiations are to be protected, and which hosts are allowed to negotiate keys. The absence of a key exchange policy is not considered an error, but without it, the IKE daemon is unable to provide dynamic key management.

A key exchange policy consists of an ordered list of key exchange rules. A key exchange rule consists of a set of security endpoints, and an action to be taken when the two security endpoints engage in an IKE phase 1 negotiation.

Optionally, a key exchange rule can contain a shared key known only to the two negotiating entities that are described in the rule. When an IKE negotiation is initiated, the current list of key exchange rules is searched for a match, based on four criteria:

- The identity of the local IKE peer, if known
- The identity of the remote IKE peer, if known
- The location (IP address) of the local IKE peer, if needed to distinguish it or if local identity is not known
- The location (IP address) of the remote IKE peer, if needed to distinguish it or if remote identity is not known

The following sample KeyExchangeRule block allows an IKE negotiation between IKE daemons at 9.2.2.2 and 9.4.4.4. A description of each line in the sample follows the sample.

```
1   KeyExchangeRule             ZoneB_KeyExRule1
2   {
3      LocalSecurityEndpoint
4      {
5         Identity              IpAddr 9.2.2.2
6         Location              9.2.2.2
7      }
8      RemoteSecurityEndpoint
9      {
10        Identity              X500dn CN=ZoneB Cert,T=IKE
ServerB,OU=endicott,O=ibm,C=US
11        Location              9.4.0.0/16
12        CaLabel               CA4endicott
```

```
13     }
14     KeyExchangeActionRef     Gold-RSA
15     SharedKey                Ascii TheEagleHasLanded
16 }
```

Line
Description

1
The KeyExchangeRule keyword, followed by a required user-defined name.

2
An open brace ({) marks the beginning of the KeyExchangeRule statement block.

3
The LocalSecurityEndpoint statement identifies a local security endpoint, or local IKE peer.

4
An open brace ({) marks the beginning of the LocalSecurityEndpoint statement block.

5
The identity of the local security endpoint that must match this rule. This can be one of five types:

- Fqdn
- IpAddr
- KeyID
- UserAtFqdn
- X500dn

In the example, an IP address is used as the identity value.

6
The IP address of the local IKE peer.

7
A close brace (}) marks the end of the LocalSecurityEndpoint statement block.

8
The RemoteSecurityEndpoint statement identifies a remote security endpoint, or remote IKE peer. The RemoteSecurityEndpoint statement can also be used to define a related group of remote IKE peers by using wildcard values for identity and location.

9
An open brace ({) marks the beginning of the RemoteSecurityEndpoint statement block.

10
The identity of the remote security endpoint that must match this rule. This can be one of five types:

- Fqdn
- IpAddr
- KeyID
- UserAtFqdn
- X500dn

In the example, an X.500 distinguished name is used as the identity value.

11
The IP subnetwork that defines a group of remote IKE peers.

12
Used only for digital signature peer authentication. Specifies the certificate authority that is advertised to the remote security endpoint as an acceptable authority. The value for this parameter must be the label of a certificate authority that is defined in RACF. The CaLabel parameter can be specified multiple times.

13
A close brace (}) marks the end of the RemoteSecurityEndpoint statement.

14
A reference to a key exchange action that has been defined elsewhere, in either the common or the stack-specific IP security configuration file, as follows:

```
KeyExchangeAction            Gold-RSA
{
   HowToInitiate              main
   HowToRespondIKEv1          main
   KeyExchangeOffer
   {
      HowToEncrypt               3DES
      HowToAuthMsgs              SHA1
      HowToAuthPeers             RsaSignature
   }
}
```

The KeyExchangeAction statement specifies the detailed parameters that govern a phase 1 negotiation between these two security endpoints, such as who can begin the negotiation and what type of encryption is used.

15
An optional shared key used only for pre-shared key host authentication.

16
A close brace (}) marks the end of the KeyExchangeRule statement block.

Example 1

The following key exchange rule for an IKEv1 Aggressive-mode phase 1 negotiation uses pre-shared key authentication:

```
KeyExchangeRule              Admin_KeyExRule1
{
  LocalSecurityEndpointRef   Internal_IKED
  RemoteSecurityEndpointRef  Admin_IKED
  KeyExchangeActionRef       Bronze-PSK
  SharedKey                  Ascii TheEagleHasLanded
}
```

This rule defines the parameters for the IKEv1 phase 1 negotiation between two hosts that are identified by the security endpoints Internal_IKED and Admin_IKED (presumed to be defined elsewhere in the policy file). The specifics of the negotiation are covered by the Bronze-PSK action as follows:

```
KeyExchangeAction        Bronze-PSK
{
   HowToInitiate         Aggressive
   HowToRespondIKEv1     Aggressive
   KeyExchangeOffer
   {
      HowToEncrypt       DES
      HowToAuthMsgs      SHA1
      HowToAuthPeers     PreSharedKey
   }
}
```

The optional SharedKey parameter is required only when the pre-shared key authentication method is used for the phase 1 negotiation.

Example 2

The following KeyExchangeRule statement for an IKEv1 Main-mode phase 1 negotiation uses digital signature authentication:

```
KeyExchangeRule              ZoneA_KeyExRule1
{
  LocalSecurityEndpointRef   Internal_IKED
  RemoteSecurityEndpointRef  ZoneA_IKED
```

```
  KeyExchangeActionRef       Silver-RSA
}
```

The referenced objects are presumed to be defined elsewhere in the policy file. This rule defines the parameters for the IKEv1 phase 1 negotiation between two hosts that are identified by the security endpoints Internal_IKED and ZoneA_IKED. The specifics of the negotiation are covered by the Silver-RSA action as follows:

```
KeyExchangeAction      Silver-RSA
{
   HowToInitiate       main
   HowToRespondIKEv1   main
   KeyExchangeOffer
   {
      HowToEncrypt     DES
      HowToAuthMsgs    SHA1
      HowToAuthPeers   RsaSignature
   }
}
```

Example 3

The following key exchange rule for an IKEv2 phase 1 negotiation uses digital signature authentication:

```
KeyExchangeRule             IKEv2_Example
{
  LocalSecurityEndpointRef  Internal_IKED
  RemoteSecurityEndpointRef ZoneA_IKED
  KeyExchangeActionRef      IKEv2-DigitalSignature
}
```

This rule defines the parameters for the IKEv2 phase 1 negotiation between two hosts that are identified by the security endpoints Internal_IKED and ZoneA_IKED (presumed to be defined elsewhere in the policy file). The specifics of the negotiation are covered by the IKEv2-DigitalSignature action as follows:

```
KeyExchangeAction            IKEv2-DigitalSignature
{
   HowToInitiate             IKEv2
   HowToAuthMe               DigitalSignature
   ReauthInterval            0
   BypassIpValidation        Yes
   KeyExchangeOffer
   {
      HowToEncrypt           AES_CBC KeyLength 128
      HowToVerifyMsgs        HMAC_SHA1_96
      PseudoRandomFunction   HMAC_SHA1
      HowToAuthPeers         RsaSignature
   }
}
```

Key exchange rule order

“Example 1” on page 953 and “Example 2” on page 953 show individual key exchange rules. A complete key exchange policy contains any number of key exchange rules. Key exchange rules in a key exchange policy are searched in the order listed. In the process of an IKE negotiation, Policy Agent searches the list of active key exchange rules to locate a best match. It is possible for more than one key exchange rule to match a pending IKE negotiation. For this reason, the list of key exchange rules in the key exchange policy should be ordered from most specific to least specific in much the same way as the IP filter rules. If the key exchange policy contains key exchange rules with both unique and wildcard security endpoints, the most specific definitions should be placed higher in the list than the wildcard definitions.

For instance, there might be one key exchange rule governing a connection from an internal administrative machine, and another key exchange rule governing all other hosts on the internal network, as follows:

```
KeyExchangeRule             Admin_KeyExRule1
{
  LocalSecurityEndpointRef  Internal_IKED
  RemoteSecurityEndpointRef Admin_IKED
  KeyExchangeActionRef      Bronze-PSK
```

```
  SharedKey                   Ascii TheEagleHasLanded
}

KeyExchangeRule               ZoneA_KeyExRule1
{
  LocalSecurityEndpointRef    Internal_IKED
  RemoteSecurityEndpointRef   ZoneA_IKED
  KeyExchangeActionRef        Silver-RSA
}
```

In this case, because the remote IKE peer that is defined by the remote security endpoint Admin_IKED matches both the Admin_KeyExRule1 and ZoneA_KeyExRule1 rules, the Admin_KeyExRule1 rule should be placed ahead of the ZoneA_KeyExRule1 rule in the key exchange policy as follows:

```
KeyExchangePolicy
{
  KeyExchangeRuleRef          Admin_KeyExRule1
  KeyExchangeRuleRef          ZoneA_KeyExRule1
  KeyExchangeRuleRef          ZoneB_KeyExRule1
  KeyExchangeRuleRef          ZoneC_KeyExRule1
}
```

Local dynamic VPN policy

A local dynamic VPN policy is required only if the IPSec negotiation is started through command-line activation using the **ipsec** command, or through automatic activation due to a local dynamic VPN policy update. IPSec negotiations can be initiated in one of four ways:

- On-demand

  The negotiation initiates when an outbound IP packet matches a filter rule with an ipsec action.
- Remotely

  The remote peer initiates the request, and the local IKE daemon responds.
- Command-line activation

  The **ipsec** command enables you to manually initiate an IPSec negotiation. (Not to be confused with manual tunnels, which do not use the IKE daemon at all.)
- Autoactivation

  The negotiation begins when either the stack or the IKE daemon initializes and both are active, or when the local dynamic VPN policy is updated.

A local dynamic VPN policy is required only in the last two cases.

A local dynamic VPN policy consists of an unordered list of local dynamic VPN rules. The negotiation for a phase 2 Security Association requires that the two communicating hosts agree on two data endpoints that the Security Association covers, the protocols the Security Association covers, and the ports that the Security Association covers. This information is then stored in the phase 2 Security Association, which is consulted each time relevant IPSec traffic needs to be encapsulated or decapsulated. The purpose of the local dynamic VPN rule is to define these requirements for each Security Association that is configured.

The following sample LocalDynVpnRule statement defines the parameters for the negotiation of a phase 2 Security Association for TN3270E Telnet server traffic between a server (9.1.1.1) and a client (9.4.4.100). A description of each line follows the sample.

```
1  LocalDynVpnRule      TelnetSA
2  {
3          LocalIP           9.1.1.1
4          RemoteIP          9.4.4.100
5          LocalDataPort     23
6          RemoteDataPort    0
7          Protocol          tcp
8          Autoactivate      yes
9  }
```

Line
Description

1
The LocalDynVpnRule keyword and user-defined name.

2
An open brace ({) marks the start of the LocalDynVpnRule statement block.

3
The local address of IP traffic that this Security Association is to protect. The address can be either a single address or a range of addresses. However, if the address is not a single address, this Security Association must be negotiated for tunnel mode.

4
The remote address of IP traffic that this Security Association is to protect. The address can be either a single address or a range of addresses. However, if the address is not a single address, this Security Association must be negotiated for tunnel mode.

5
The local ports for IP traffic that this Security Association is to protect. The port must be either a single port or all ports. A range of ports is not allowed. A value of 0 indicates all ports.

6
The remote ports for IP traffic that this Security Association is to protect. The port must be either a single port or all ports. A range of ports is not allowed. A value of 0 indicates all ports.

7
The protocol that this Security Association is to protect. This value can be numeric. It must define a single protocol or all protocols.

8
Indicates that this Security Association is to be activated when the stack and IKE daemon are active. No user intervention is required.

9
A close brace (}) marks the end of the LocalDynVpnRule statement block.

An on-demand Security Association does not require a local dynamic VPN rule definition. All of the parameters for the negotiation of an on-demand phase 2 Security Association can be inferred from one of two places, either the packet that began the on-demand activation or the filter rule on which the packet matched. The packet always provides a single value for address, protocol, port, type, and code. The filter rule, however, can allow for a range of values for IP address, protocol, port, type, or code. The granularity setting of the IpLocalStartAction statement determines whether the information is taken from the packet or from the matching filter rule. For more information regarding the IpLocalStartAction statement, see z/OS Communications Server: IP Configuration Reference.

Security Associations can be defined as wide or narrow with respect to IP addresses or ports and protocols. If a Security Association is wide with respect to IP address, the same Security Association is used to protect data between multiple endpoints. If a Security Association is wide with respect to ports and protocols, the same Security Association is used to protect multiple traffic types. Conversely, a narrowly defined Security Association can be used to protect specific data endpoints (based on IP address) or specific traffic types (based on commonly used ports and protocols for network services). Security Associations negotiated with IKEv2 can also be narrow with respect to types and codes.

Example 1 - wide Security Association

The following rule allows any type of traffic to flow between PublicServerAddressA1 and SubnetC using the same Security Association. PublicServerAddressA1 and SubnetC can be defined in either the common or the stack-specific IP security configuration file. The AutoActivate parameter causes the IKE negotiation to initiate when the stack or IKE initializes.

```
LocalDynVpnRule       ZoneC_VPN-All-traffic
{
     LocalIpRef          PublicServerAddressA1
     RemoteIpSetRef      SubnetC
     Protocol            all
     AutoActivate        yes
}
```

```
IpAddr                PublicServerAddressA1
{
  Addr                9.3.3.3
}

IpAddrSet             SubnetC
{
  Prefix              9.6.0.0/16
}
```

Example 2 - narrow Security Association

If narrow Security Associations are used for IPSec-protected FTP traffic, two VPN definitions are required, one for the data connection and one for the control connection. The following rules are from the server's perspective. The FTP client connecting from BranchOfficeAddressC1 to the PublicServerAddressA1 ports 20 and 21 uses the respective ZoneC FTP VPNs.

```
LocalDynVpnRule              ZoneC_VPN-FTP-Data
{
     LocalIpRef                 PublicServerAddressA1
     RemoteIpRef                BranchOfficeAddressC1
     LocalDataPort              20
     RemoteDataPort             0
     Protocol                   tcp
}

LocalDynVpnRule              ZoneC_VPN-FTP-Control
{
     LocalIpRef                 PublicServerAddressA1
     RemoteIpRef                BranchOfficeAddressC1
     LocalDataPort              21
     RemoteDataPort             0
     Protocol                   tcp
}

IpAddr                       PublicServerAddressA1
{
  Addr                       9.3.3.3
}

IpAddr                       BranchOfficeAddressC1
{
  Addr                       9.5.5.5
}
```

Quick start using IP filtering and IPSec host-to-host

The following sample shows a complete IP security policy allowing connections from a secure server (9.1.1.1) to an administrative machine (9.1.1.2) on an internal network. It represents the absolute minimum number of items that need to be configured for the IKED to provide IPSec protection with dynamic key management between two hosts.

Tip: You can modify the quick start sample for IPv6 by replacing all IPv4 addresses with IPv6 addresses.

This IP security policy allows IKE negotiations in the clear (UDP, port 500 traffic), while authenticating and encrypting all other traffic using the ESP IPSec protocol. The policy relies almost exclusively on the z/OS IP security policy defaults, including MD5 and DES for the phase 1 Security Association and ESP/MD5 ESP/DES for the phase 2 Security Association. For a complete description of IP security policy configuration statements and their defaults, see z/OS Communications Server: IP Configuration Reference.

```
#------------------------------------------------------
# Quick-Start IP Security policy
#------------------------------------------------------
IpFilterPolicy
{
  PreDecap                     off
  FilterLogging                on
  AllowOnDemand                yes

  IpFilterRule                 QuickStartRule1
  {
```

```
    IpSourceAddr               9.1.1.1
    IpDestAddr                 9.1.1.2
    IpService
    {
      SourcePortRange          500
      DestinationPortRange     500
      Protocol                 udp
      Direction                bidirectional
      Routing                  local
    }
    IpGenericFilterActionRef   permit
  }

  IpFilterRule                 QuickStartRule2
  {
    IpSourceAddr               9.1.1.1
    IpDestAddr                 9.1.1.2
    IpService
    {
      Direction                bidirectional
      Routing                  local
    }
    IpGenericFilterActionRef   ipsec
    IpDynVpnActionRef          TransportMode
  }
}

KeyExchangePolicy
{
  KeyExchangeRule              QuickStart_KeyExRule
  {
     LocalSecurityEndpoint
     {
       Identity                IpAddr 9.1.1.1
       Location                9.1.1.1
     }
     RemoteSecurityEndpoint
     {
       Identity                IpAddr 9.1.1.2
       Location                9.1.1.2
     }
    KeyExchangeActionRef       QuickStart_KeyExAction
    SharedKey                  Ascii TheEagleHasLanded
  }
}

#-------------------------------------------------------
# Reusable actions
#-------------------------------------------------------
IpGenericFilterAction    permit
{
  IpFilterAction         permit
}

IpGenericFilterAction    ipsec
{
  IpFilterAction         ipsec
  IpFilterLogging        yes LogDeny
}

KeyExchangeAction        QuickStart_KeyExAction
{
  KeyExchangeOffer
  {
    HowToAuthPeers       PreSharedKey
  }
}

IpDynVpnAction           TransportMode
{
  IpDataOffer
  {
    HowToEncap           transport
  }
}
```

For all IKE negotiations, there must be a corresponding and consistent configuration on the remote host. In this case, if the remote system is running with z/OS IP security, the corresponding policy for the remote system can be generated merely by transposing all instances of local and remote IP addresses.

Displaying filters, rules, and actions

To display the filter rules for the quick start policy after they have been installed in the stack, enter the following UNIX System Services command:

```
ipsec -f display -r detail -c current

CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 10:29:51 2012
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Policy    Scope:    Current            TotAvail: 8
Logging:  On              Predecap: Off                DVIPSec:  No
NatKeepAlive:  0          FIPS140:  No
Defensive Mode: Inactive

FilterName:                   QuickStartRule1
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     UDP(17)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                9.1.1.1
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   500
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.2
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     500
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
********************************************************************************
FilterName:                   QuickStartRule1
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
```

```
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     UDP(17)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                9.1.1.2
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   500
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.1
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     500
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   QuickStartRule2
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                TransportMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     Yes
SecurityClass:                0
Logging:                      Deny
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.1.1.1
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.2
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
```

```
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   QuickStartRule2
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                TransportMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     Yes
SecurityClass:                0
Logging:                      Deny
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.1.1.2
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.1
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   DenyAllRule_Generated__________Inbnd
FilterNameExtension:          n/a
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
```

```
Scope:                          Both
Direction:                      Inbound
OnDemand:                       n/a
SecurityClass:                  0
Logging:                        None
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:28:42
UpdateTime:                     2012/02/14 10:28:42
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  34
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     DenyAllRule_Generated__________Outbnd
FilterNameExtension:            n/a
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
DefensiveType:                  n/a
State:                          Active
Action:                         Deny
Scope:                          Both
Direction:                      Outbound
OnDemand:                       n/a
SecurityClass:                  0
Logging:                        None
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
```

```
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                7
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   DenyAllRule_Generated__________Inbnd_v6
FilterNameExtension:          n/a
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
Action:                       Deny
Scope:                        Both
Direction:                    Inbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                1
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   DenyAllRule_Generated__________Outbnd_v6
FilterNameExtension:          n/a
GroupName:                    n/a
LocalStartActionName:         n/a
VpnActionName:                n/a
TunnelID:                     0x00
Type:                         Generic
DefensiveType:                n/a
State:                        Active
```

```
Action:                       Deny
Scope:                        Both
Direction:                    Outbound
OnDemand:                     n/a
SecurityClass:                0
Logging:                      None
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                ::
SourceAddressPrefix:          0
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  ::
DestAddressPrefix:            0
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:28:42
UpdateTime:                   2012/02/14 10:28:42
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*******************************************************************************

8 entries selected
```

Each IP service in the example uses the bidirectional keyword. Therefore, two rules are created for each IP service, one outbound and one inbound. When the IP filter rules are expanded in this way, the specific filter rules are distinguished from each other by a unique numeric value in the FilterNameExtension field.

Note that the last four deny rules are not explicitly coded in the IP security configuration file, but are added by the system in keeping with a default-deny policy.

For more information on displaying active filters with the **ipsec** command, see "Displaying active filters with the ipsec command" on page 1051.

To view the quick start filter rules using the **pasearch** command, issue the following command:

```
pasearch -v f

TCP/IP pasearch CS V1R12                  Image Name: TCPCS
  Date:                 02/16/2010        Time:  10:30:47
  IPSec Instance Id:    1266334122

policyRule:             QuickStartRule1
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               106               ForLoadDist:       False
  Priority:             6                 Sequence Actions:  Don't Care
  No. Policy Action:    1                 ConditionListType: CNF
  IpSecType:            policyIpFilter
  policyAction:         permit
   ActionType:          IpFilter GenericFilter
   Action Sequence:     0
  Time Periods:
```

```
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:               NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  Condition Work Level:      0
    Group Number:       0                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:          NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  IpSec Condition Work:                  NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          9.1.1.1
     ToAddr:            9.1.1.1
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  Condition Work Level:      1
    Group Number:       1                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:          NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  IpSec Condition Work:                  NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
     FromAddr:          9.1.1.2
     ToAddr:            9.1.1.2
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  Condition Work Level:      2
    Group Number:       3                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:          NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  IpSec Condition Work:                  NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
```

```
     Protocol:          UDP  (17)
      SrcPortFrom:      500               SrcPortTo:         500
      DestPortFrom:     500               DestPortTo:        500
     Direction:         Bidirectional
     RouteType:         Local             SecurityClass:     0
     FragmentsOnly:     No
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010

  IpFilter Action:      permit
    Version:            3                 Status:            Active
    Scope:              GenericFilter
    ipFilterAction:     Permit            IpFilterLogging:   No
    DiscardAction:      Silent
    Policy created: Tue Feb 16 10:28:42 2010
    Policy updated: Tue Feb 16 10:28:42 2010

policyRule:             QuickStartRule2
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               105               ForLoadDist:       False
  Priority:             5                 Sequence Actions:  Don't Care
  No. Policy Action:    2                 ConditionListType: CNF
  IpSecType:            policyIpFilter
  policyAction:         ipsec
   ActionType:          IpFilter GenericFilter
   Action Sequence:     0
  policyAction:         TransportMode
   ActionType:          IpFilter DynamicVpn
   Action Sequence:     0
  Time Periods:
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:                NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  Condition Work Level:      0
    Group Number:       0                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:           NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  IpSec Condition Work:                   NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          9.1.1.1
     ToAddr:            9.1.1.1
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  Condition Work Level:      1
    Group Number:       1                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:           NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
```

```
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  IpSec Condition Work:                   NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
     FromAddr:          9.1.1.2
     ToAddr:            9.1.1.2
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  Condition Work Level:      2
    Group Number:       3                 Cond Count:        2
    Ignore:             No
  IpSec Condition Work Summary:           NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          0
     Direction:         0
     RouteType:         0                 SecurityClass:     0
     FragmentsOnly:     No
  IpSec Condition Work:                   NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
    Destination Address:
    Service Condition:
     Protocol:          All
     Direction:         Bidirectional
     RouteType:         Local             SecurityClass:     0
     FragmentsOnly:     No
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010

  IpFilter Action:      ipsec
    Version:            3                 Status:            Active
    Scope:              GenericFilter
    ipFilterAction:     IPSec             IpFilterLogging:   Yes Logdeny
    DiscardAction:      Silent
    Policy created: Tue Feb 16 10:28:42 2010
    Policy updated: Tue Feb 16 10:28:42 2010

  IpFilter Action:      TransportMode
    Version:            3                 Status:            Active
    Scope:              DynamicVpn
    Initiation:         Either            VpnLife:           1440
    AcceptablePfs:      None
    InitiateWithPfs:    None              IpDataOfferNum:    1
    PassthroughDSCP:    Yes               PassthroughDF:     Yes
    HowToEncapIKEv2:    Either
    IPDataOffer:        0
     HowToEncap:        Transport
     HowToEncrypt:      DES               KeyLength:         N/A
     HowToAuth:         ESP               HowToAuthAlgr:     HMAC_MD5
     RefLifeTmPropose:  240
     RefLifeTmAcptMin:  120               RefLifeTmAcptMax:  480
     RefLifeSzPropose:  None
     RefLifeSzAccept :  None
    Policy created: Tue Feb 16 10:28:42 2010
    Policy updated: Tue Feb 16 10:28:42 2010

policyRule:             DenyAllRule_Generated__________Inbnd
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               104               ForLoadDist:       False
  Priority:             4                 Sequence Actions:  Don't Care
  No. Policy Action:    0
  IpSecType:            policyIpFilter
  Time Periods:
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
```

```
   TimeZone:            Local
  IpSec Condition Summary:               NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          All4
     ToAddr:            All4
    Destination Address:
     FromAddr:          All4
     ToAddr:            All4
    Service Condition:
     Protocol:          All
     Direction:         Inbound
     RouteType:         Either            SecurityClass:     0
     FragmentsOnly:     No
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010

policyRule:             DenyAllRule_Generated__________Outbnd
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               103               ForLoadDist:       False
  Priority:             3                 Sequence Actions:  Don't Care
  No. Policy Action:    0
  IpSecType:            policyIpFilter
  Time Periods:
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:               NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          All4
     ToAddr:            All4
    Destination Address:
     FromAddr:          All4
     ToAddr:            All4
    Service Condition:
     Protocol:          All
     Direction:         Outbound
     RouteType:         Either            SecurityClass:     0
     FragmentsOnly:     No
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010

policyRule:             DenyAllRule_Generated__________Inbnd_v6
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               102               ForLoadDist:       False
  Priority:             2                 Sequence Actions:  Don't Care
  No. Policy Action:    0
  IpSecType:            policyIpFilter
  Time Periods:
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:               NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          All6
     ToAddr:            All6
    Destination Address:
     FromAddr:          All6
     ToAddr:            All6
    Service Condition:
     Protocol:          All
     Direction:         Inbound
     RouteType:         Either            SecurityClass:     0
     FragmentsOnly:     No
```

```
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010

policyRule:             DenyAllRule_Generated__________Outbnd_v6
  Rule Type:            IpFilter
  Version:              3                 Status:            Active
  Weight:               101               ForLoadDist:       False
  Priority:             1                 Sequence Actions:  Don't Care
  No. Policy Action:    0
  IpSecType:            policyIpFilter
  Time Periods:
   Day of Month Mask:
   First to Last:       11111111111111111111111111111111
   Last to First:       11111111111111111111111111111111
   Month of Yr Mask:    111111111111
   Day of Week Mask:    1111111  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      24:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:                NegativeIndicator: Off
   IpFilter Condition:
    Source Address:
     FromAddr:          All6
     ToAddr:            All6
    Destination Address:
     FromAddr:          All6
     ToAddr:            All6
    Service Condition:
     Protocol:          All
     Direction:         Outbound
     RouteType:         Either            SecurityClass:     0
     FragmentsOnly:     No
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010
```

For more information on displaying filter rules with the **pasearch** command, see "Displaying filter rules with the pasearch command" on page 1069.

To display the key exchange rules and actions for the quick start IP security policy, issue the following command:

```
pasearch -v k

TCP/IP pasearch CS V1R12                Image Name: TCPCS
  Date:                 02/16/2010        Time:  10:31:07
  IPSec Instance Id:    1266334122

policyRule:             QuickStart_KeyExRule
  Rule Type:            KeyExchange
  Version:              3                 Status:            Active
  Weight:               101               ForLoadDist:       False
  Priority:             1                 Sequence Actions:  Don't Care
  No. Policy Action:    1
  IpSecType:            policyKeyExchange
  policyAction:         QuickStart_KeyExAction
   ActionType:          KeyExchange
   Action Sequence:     0
  Time Periods:
   Day of Month Mask:   00000000000000000000000000000000
   Month of Yr Mask:    000000000000
   Day of Week Mask:    0000000  (Sunday - Saturday)
   Start Date Time:     None
   End Date Time:       None
   Fr TimeOfDay:        00:00             To TimeOfDay:      00:00
   Fr TimeOfDay UTC:    00:00             To TimeOfDay UTC:  00:00
   TimeZone:            Local
  IpSec Condition Summary:                NegativeIndicator: Off
   KeyExchange Condition:
    LocalSecurityEndPoint:
     Location:
      FromAddr:         9.1.1.1
      ToAddr:           9.1.1.1
     Identity:
      IpAddr:
       FromAddr:        9.1.1.1
       ToAddr:          9.1.1.1
    RemoteSecurityEndPoint:
```

```
    Location:
     FromAddr:         9.1.1.2
     ToAddr:           9.1.1.2
    Identity:
     IpAddr:
      FromAddr:        9.1.1.2
      ToAddr:          9.1.1.2
  Policy created: Tue Feb 16 10:28:42 2010
  Policy updated: Tue Feb 16 10:28:42 2010

  KeyExchange Action:   QuickStart_KeyExAction
    Version:            3                 Status:            Active
    HowToInitiate:      Main              HowToRespondIKEv1: Either
    AllowNat:           No                FilterByIdentity:  No
    HowToAuthMe:        DigitalSignature  ReauthInterval:    0
    BypassIpValidation: No                CertURLLookupPref: Tolerate
    RevocationChecking: Loose
    KeyExchangeOffer:   0
     HowToEncrypt:      DES               KeyLength:         N/A
     HowToAuthPeers:    PresharedKey      DHGroup:           Group2
     HowToAuthMsgs:     MD5
     HowToVerifyMsgs:   HMAC_SHA1_96      PseudoRandomFunc:  HMAC_SHA1
     RefLifeTmPropose:  480
     RefLifeTmAcptMin:  240               RefLifeTmAcptMax:  1440
     RefLifeSzPropose:  None
     RefLifeSzAccept :  None
    Policy created: Tue Feb 16 10:28:42 2010
    Policy updated: Tue Feb 16 10:28:42 2010
```

Activating the quick start Security Association

The quick start policy enables on-demand activation of a Security Association between the two endpoints. The Security Association is a wide Security Association allowing any type of traffic. Therefore, it can be activated by sending any type of traffic from the local host at 9.1.1.1 to the remote host at 9.1.1.2 as follows:

```
ping -s 9.1.1.1 9.1.1.2
CS V2R5: Pinging host 9.1.1.2
sendto(): EDC5111I Permission denied. (errno2=0x74420291)
```

Because the Security Association does not exist, the initial **ping** attempt fails. After the negotiation for the Security Association has activated and the Security Association is established, a subsequent **ping** attempt succeeds as follows:

```
ping -s 9.1.1.1 9.1.1.2
CS V2R5: Pinging host 9.1.1.2
Ping #1 response took 0.001 seconds.(1.378 milliseconds)
```

Displaying the quick start Security Associations

Use the **ipsec** command to display both the phase 1 and phase 2 Security Associations between 9.1.1.1 and 9.1.1.2. The following command displays the phase 1 Security Associations:

```
ipsec -k display -r detail

CS V2R5 ipsec  Stack Name: TCPCS  Tue Feb 16 10:38:12 2020
Primary:  IKE tunnel      Function: Display            Format:   Detail
Source:   IKED            Scope:    Current            TotAvail: n/a

TunnelID:                     K1
Generation:                   1
IKEVersion:                   1.0
KeyExchangeRuleName:          QuickStart_KeyExRule
KeyExchangeActionName:        QuickStart_KeyExAction
LocalEndPoint:                9.1.1.1
LocalIDType:                  ID_IPV4_ADDR
LocalID:                      9.1.1.1
RemoteEndPoint:               9.1.1.2
RemoteIDType:                 ID_IPV4_ADDR
RemoteID:                     9.1.1.2
ExchangeMode:                 Main
State:                        DONE
AuthenticationAlgorithm:      HMAC-MD5
```

```
EncryptionAlgorithm:          DES-CBC
 KeyLength:                   n/a
PseudoRandomFunction:         HMAC-MD5
DiffieHellmanGroup:           2
LocalAuthenticationMethod:    PresharedKey
RemoteAuthenticationMethod:   PresharedKey
InitiatorCookie:              0x7456F943AA0154BB
ResponderCookie:              0xA344ED85C5D00154
Lifesize:                     0K
CurrentByteCount:             288b
Lifetime:                     480m
LifetimeRefresh:              2020/02/16 18:26:45
LifetimeExpires:              2020/02/16 18:37:43
ReauthInterval:               480m
ReauthTime:                   2020/02/16 18:26:45
Role:                         Initiator
AssociatedDynamicTunnels:     1
NATTSupportLevel:             None
NATInFrntLclScEndPnt:         No
NATInFrntRmtScEndPnt:         No
zOSCanInitiateP1SA:           Yes
AllowNat:                     No
RmtNAPTDetected:              No
RmtUdpEncapPort:              n/a
LocalCertExpires:             n/a
LocalSerialNumber:            n/a
LocalIssuerDNLength:          0
LocalIssuerDN:                n/a
LocalSubjectDNLength:         0
LocalSubjectDN:               n/a
RemoteCertExpires:            n/a
RemoteSerialNumber:           n/a
RemoteIssuerDNLength:         0
RemoteIssuerDN:               n/a
RemoteSubjectDNLength:        0
RemoteSubjectDN:              n/a
*************************************************************************

1 entries selected
```

In addition to information relating specifically to the phase 2 Security Association, use the **ipsec -y display** command to find the phase 1 that protects it. The ParentIKETunnelID field shows the associated phase 1, which is the same as the TunnelID from the previous **ipsec -k display** command.

```
ipsec -y display -r detail

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 10:39:25 2010
Primary:  Dynamic tunnel  Function: Display            Format:   Detail
Source:   Stack           Scope:    Current            TotAvail: 1

TunnelID:                     Y2
Generation:                   1
IKEVersion:                   1.0
ParentIKETunnelID:            K1
VpnActionName:                TransportMode
LocalDynVpnRule:              n/a
State:                        Active
HowToEncap:                   Transport
LocalEndPoint:                9.1.1.1
RemoteEndPoint:               9.1.1.2
LocalAddressBase:             9.1.1.1
LocalAddressPrefix:           n/a
LocalAddressRange:            n/a
RemoteAddressBase:            9.1.1.2
RemoteAddressPrefix:          n/a
RemoteAddressRange:           n/a
HowToAuth:                    ESP
 AuthAlgorithm:               HMAC-MD5
 AuthInboundSpi:              1878088104 (0x6FF159A8)
 AuthOutboundSpi:             270783814  (0x1023D546)
HowToEncrypt:                 DES-CBC
 KeyLength:                   n/a
 EncryptInboundSpi:           1878088104 (0x6FF159A8)
 EncryptOutboundSpi:          270783814  (0x1023D546)
Protocol:                     ALL(0)
LocalPort:                    n/a
LocalPortRange:               n/a
RemotePort:                   n/a
```

```
RemotePortRange:              n/a
Type:                         n/a
TypeRange:                    n/a
Code:                         n/a
CodeRange:                    n/a
OutboundPackets:              1
OutboundBytes:                264
InboundPackets:               1
InboundBytes:                 264
Lifesize:                     0K
LifesizeRefresh:              0K
CurrentByteCount:             0b
LifetimeRefresh:              2010/02/16 14:26:22
LifetimeExpires:              2010/02/16 14:37:43
CurrentTime:                  2010/02/16 10:39:25
VPNLifeExpires:               2010/02/17 10:37:43
NAT Traversal Topology:
  UdpEncapMode:               No
  LclNATDetected:             No
  RmtNATDetected:             No
  RmtNAPTDetected:            No
  RmtIsGw:                    n/a
  RmtIsZOS:                   n/a
  zOSCanInitP2SA:             n/a
  RmtUdpEncapPort:            n/a
  SrcNATOARcvd:               n/a
  DstNATOARcvd:               n/a
PassthroughDF:                n/a
PassthroughDSCP:              n/a
*********************************************************************

1 entries selected
```

Steps for configuring IP security policy

These steps are used to configure the specific security models.

Procedure

Perform the following steps to configure IP security policy.

1. Determine the number of zones to be protected. A zone typically equates to a subnetwork that is reachable by the host server, but can be any group of IP addresses that are conceptually related. The internal network can be defined in one or several zones, while any external network or group of networks can be placed in separate zones. Each zone should have meaning related to the site's security policy. If a zone maps to a physical interface, optionally assign a security class to all interfaces in that zone.
2. For each zone, determine what services are allowed and define an IpService statement for each wanted service. Services are defined by the protocols and well-known ports that they use.
3. Determine the data endpoints to be protected. Typically, this is both a local and remote IP address, or subnetwork.
4. Determine what level of security is needed between each set of data endpoints. The level of security can be deny, permit, or ipsec.
5. Configure an IpGenericFilterAction statement for the level of security that is required (permit, deny, ipsec), including whether the connection should be logged.
6. If IPSec is required between any two endpoints:
   a) Configure a KeyExchangePolicy statement that defines the parameters of the phase 1 negotiation:
      i) Determine the required type and strength of protection for the phase 1 Security Association.
      ii) Decide what type of peer authentication will be used:
         - If digital signature, set up RACF certificates and certificate authority information.
         - If pre-shared key, create a secret key that is known to both peers.
      iii) Decide whether NAT traversal will be allowed. If the network topology contains one or more NAT devices that must be traversed by the phase 1 Security Association, NAT traversal should be allowed.

iv) Configure a KeyExchangeOffer statement.

v) Determine negotiation mode, IKEv1 Main, IKEv1 Aggressive, or IKEv2.

vi) Configure a KeyExchangeAction statement.

vii) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement.

viii) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action.

ix) Include the KeyExchangeRule statement in the KeyExchangePolicy statement block.

b) Configure an IpDynVpnAction statement that defines the control of the phase 2 negotiation:

i) Determine the required type and strength of IPSec protection for the phase 2 Security Association.

ii) Determine whether tunnel or transport mode is required. For an IKEv2 negotiation, the appropriate mode is chosen based on topology.

iii) Configure an IpDataOffer statement that defines the parameters of the phase 2 negotiation.

iv) Determine which peer should be allowed to initiate the negotiation.

c) Decide how the Security Association is to be activated:

i) Configure an optional LocalDynVpnPolicy statement, if command-line activated or autoactivated.

ii) Configure an optional IpLocalStartAction statement, if the Security Association is to be activated locally (that is, on-demand, command-line, or autoactivation) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host Security Association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

iii) Either create an IpFilterRule statement that allows IPSec traffic (AH and ESP), or set the global PreDecap parameter of the IpFilterPolicy statement to off.

7. Define an IpFilterRule statement for each set of data endpoints.

   The rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500). If NAT traversal is allowed, create an IpFilterRule statement that allows IKE UDP traffic on port 4500.

   Rule: To allow IKE negotiations for Security Associations, IKE traffic (port 500, and optionally port 4500 for NAT traversal) must be permitted in the clear.

8. Define an IpFilterGroup statement for each zone and include the IpFilterRule statements that belong to that zone.
9. Include the IpFilterRule statements in the IpFilterPolicy block.
10. Include all configured statements in the IP security configuration file.

    For more information, see:

    - "Steps for configuring local IP security policy using only a common IP security configuration file" on page 942
    - "Steps for configuring remote IP security policy using only a common IP security configuration file" on page 942
    - "Steps for configuring local IP security policy using only a stack-specific IP security configuration file" on page 943
    - "Steps for configuring remote IP security policy using only a stack-specific IP security configuration file" on page 944
    - "Steps for configuring local IP security policy using both a stack-specific file and a common file" on page 945
    - "Steps for configuring remote IP security policy using both a stack-specific file and a common file" on page 945

Results

For examples of the use of these steps, see "Configuring specific security models" on page 974.

Configuring specific security models

Setting up IP security configuration files can be a complex task, as there are many powerful features, options, and controls. However, after the security needs of the business are identified, implementing an IP security policy becomes a matter of translating the requirements to a Policy Agent configuration file.

The choice of protection model primarily depends on the network topology. Although it is perfectly permissible to follow a single model when configuring IP security policy, the z/OS IP security function enables any number of models to be installed concurrently. Commonly, one set of rules governs internal network traffic, another protects traffic from connected networks, and a third provides security for traffic that is routed over the Internet. The following scenarios presume that you are configuring a secure server that is a multihomed host that is connected to an internal, an external, and a wide-area network that traverses the Internet. The configuration guidelines that are presented in the following subtopics are based on three business models:

- Trusted internal network (permit, deny)

  In the trusted internal network model, the server is protecting traffic that originates from hosts inside a privately controlled network. IP packets on the internal network are not generally subject to the stringent restrictions that are placed on traffic that is generated from outside the business. This model is usually more tolerant, given that users inside the company need access to internal network resources and services, such as the web, FTP, and Telnet.
- Partner company (permit, deny, strong IPSec protection)

  The partner company model consists of two interconnected networks, with the server protecting traffic that originates from hosts outside the internal network. Typically, two separate networks are physically connected to the z/OS server. Because the traffic is not restricted to internal hosts, security is usually somewhat tighter than in the trusted internal network model. Each partner company has no physical control over the machine of the other partner company. The services that are provided are determined by the needs of the business, but typically include many of the same services that are provided to the internal network, such as access to a web server, FTP, and Telnet. Though many services might be allowed between partner companies, the need for confidentiality and authentication of data is more stringent than in the trusted internal network model, because there is little to no control over the other network. IPSec is often specified to authenticate and optionally encrypt data that flows between the two networks.
- Branch office (permit, deny, strong tunnel-mode IPSec protection)

  The branch office model consists of two networks whose IP connectivity relies on the Internet. The server is protecting traffic that originates from hosts outside the internal network, which at some point is routed over the Internet. Because there is no control over any data that traverses the Internet, the need for security is greatest in this model. The services that are provided are based on business need, but typically include a subset of what is available internally. All traffic that traverses the Internet carrying vital information should be secured using some form of authentication and encryption.

Figure 123 on page 975 shows a sample network for all three security models.

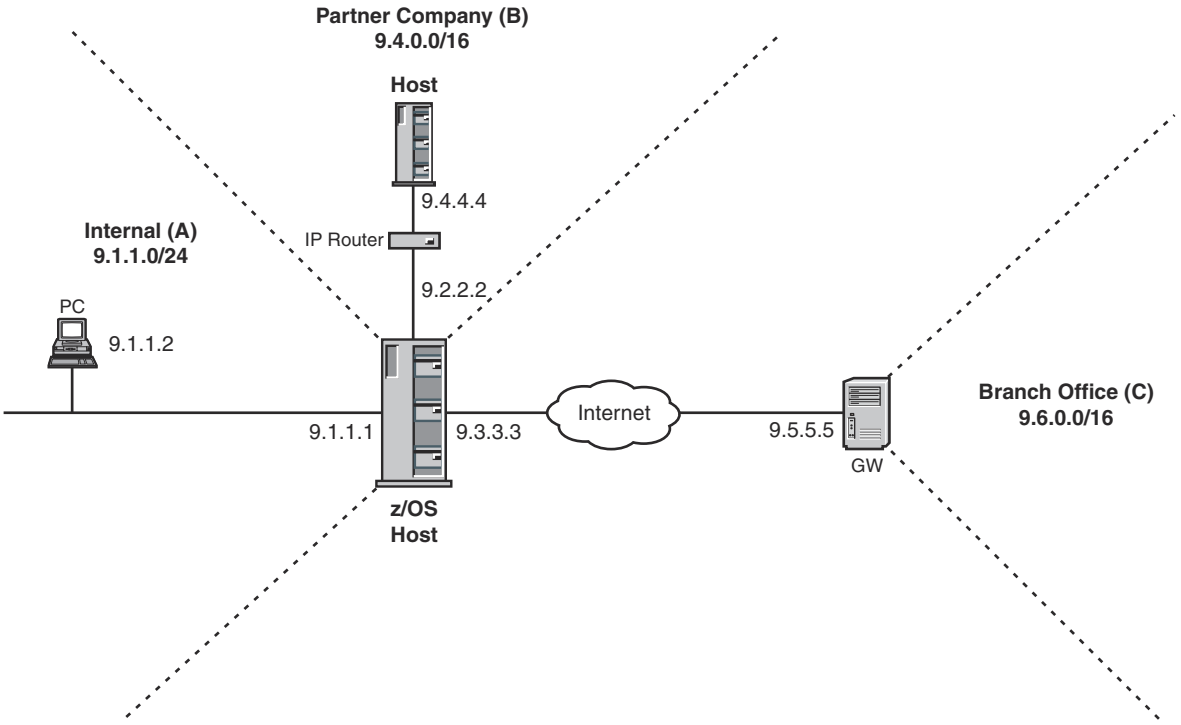


Figure 123. Security model network

The following subtopics describe how to configure these models using the steps described in "Steps for configuring IP security policy" on page 972. The policy examples assume that a default-deny policy is in place. Any traffic not explicitly permitted is blocked.

Steps for configuring the trusted internal network model (simple IP filtering)

In the trusted internal network model, the server is protecting traffic that originates from hosts inside a privately controlled network.

Before you begin

The following statements, concepts, and files are covered in the discussion of this model:

- IpFilterRule
- IpService
- IpGenericFilterAction
- References
- Groups
- Stack-specific IP security configuration file
- Common IP security configuration file

Figure 124 on page 976 shows the trusted internal network portion of the security model network.

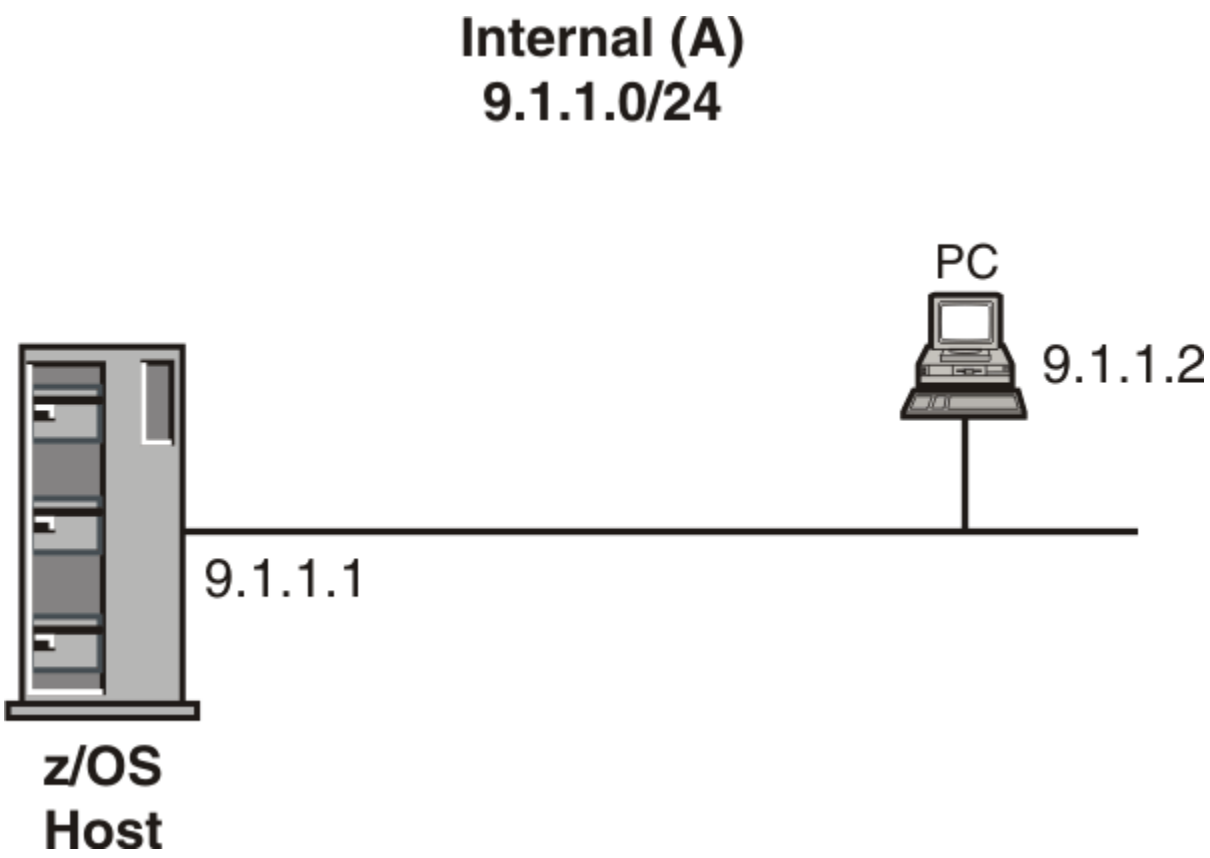


Figure 124. Trusted internal network model

For this example, assume that the following requirements must be met to control traffic on the internal network:

- Give FTP access to an administrator from a host inside the internal network (9.1.1.2) to the secure server (9.1.1.1). Log any traffic that matches this traffic pattern.
- Allow web access (HTTP, port 80) to the secure server from any host on the internal network (9.1.1.0/24). Do not log any traffic that matches this traffic pattern.
- Deny all other traffic.

Procedure

Perform the following steps to meet these requirements and configure the trusted internal network model.

1. Determine the number of zones to be protected.

   There is only one zone for this example, the internal network 9.1.1.0/24.

2. For each zone, determine what services are allowed and define an IpService block for each wanted service. Optionally, assign a security class to all interfaces in each zone.

   There are two services stated in the example requirements, HTTP and FTP. The traffic is local to this host and, therefore, the routing is designated as local. No forwarding of these services is allowed.

   Because the entire internal network is defined in one zone, you can define a unique security class for the interface with address 9.1.1.1. For this example, the SECCLASS parameter of all internal network interfaces is assigned the arbitrary value of 1, which can be interpreted to mean a trusted network. If you specify the SecurityClass parameter in the IpService block, the related interface must be assigned the same value on the SECCLASS parameter of the LINK or INTERFACE statement in

the TCP/IP profile. In this example, the traffic is allowed only over an interface with a SECCLASS parameter value of 1, presumed to be the interface connected to the internal network.

```
IpService
{
  SourcePortRange       80
  DestinationPortRange  1024 65535
  Protocol              tcp
  Direction             bidirectional InboundConnect
  Routing               local
  SecurityClass         1
}
```

Because normal FTP uses two well-known ports, two services are required, one for the control connection and one for the data connection:

```
IpService
{
  SourcePortRange       21
  DestinationPortRange  1024 65535
  Protocol              tcp
  Direction             bidirectional InboundConnect
  Routing               local
  SecurityClass         1

}

IpService
{
  SourcePortRange       20
  DestinationPortRange  1024 65535
  Protocol              tcp
  Direction             bidirectional OutboundConnect
  Routing               local
  SecurityClass         1

}
```

The InboundConnect keyword is used for services that are not allowed to initiate a TCP connection. The OutboundConnect keyword is used for services that are not allowed to receive a TCP connection request. If neither keyword is specified, either side can initiate a TCP connection.

3. Determine the data endpoints to be protected.

   There are two sets of data endpoints to be protected in this example, representing the connection from the administrative machine, and all the other hosts on the subnetwork:

```
Local  Address of secure server: 9.1.1.1
Remote Address of administrative machine: 9.1.1.2

Local  Address of secure server: 9.1.1.1
Remote Address of all hosts on internal network: 9.1.1.0/24
```

4. Determine what level of security is needed between each set of data endpoints.

   In this example, only permit is required. Therefore, no IPSec information is needed. Because z/OS IP security policy implicitly provides a default-deny policy, all other traffic is denied.

5. Configure an IpGenericFilterAction statement for the level of security (permit, deny, ipsec) that is required, including whether the connection is logged.

   Because the example requirement is to permit two types of traffic with different logging requirements, two actions are needed as follows:

```
IpGenericFilterAction    permit-log
{
   IpFilterAction       permit
   IpFilterLogging       yes
}

IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}
```

6. If IPSec is required between any two endpoints, configure a KeyExchangePolicy statement that defines the parameters of the phase 1 negotiation, configure an IpDynVpnAction statement that defines the control of the phase 2 negotiation, and decide how the Security Association is to be activated.

   IPSec is not required in this example. If there was sensitive data flowing through the internal network that needed to be confidential, IPSec could be specified to encrypt some IP packets, thereby effectively securing information that travels between two hosts on the internal network.

7. Define an IpFilterRule block for each set of data endpoints.

   Each rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500). If NAT traversal is allowed, create an IpFilterRule statement that allows IKE UDP traffic on port 4500.

   In this example, the source address refers to an address on the secure host. The destination address refers to remote hosts. The IpService statements are the ones defined in step “2” on page 976. Note that the IpGenericFilterAction statement must reference a previously defined action.

```
IpFilterRule             AdminFTP
{
   IpSourceAddr          9.1.1.1
   IpDestAddr            9.1.1.2
   IpService
   {
      SourcePortRange       21
      DestinationPortRange  1024 65535
      Protocol              tcp
      Direction             bidirectional InboundConnect
      Routing               local
      SecurityClass         1

   }
   IpService
   {
      SourcePortRange       20
      DestinationPortRange  1024 65535
      Protocol              tcp
      Direction             bidirectional OutboundConnect
      Routing               local
      SecurityClass         1
   }
   IpGenericFilterActionRef  permit-log
}
IpFilterRule             InternalNetWeb
{
   IpSourceAddr          9.1.1.1
   IpDestAddrSet         9.1.1.0/24
   IpService
   {
      SourcePortRange       80
      DestinationPortRange  1024 65535
      Protocol              tcp
      Direction             bidirectional InboundConnect
      Routing               local
      SecurityClass         1
   }
   IpGenericFilterActionRef  permit-nolog
}

IpGenericFilterAction        permit-log
{
  IpFilterAction             permit
  IpFilterLogging            yes
}

IpGenericFilterAction        permit-nolog
{
  IpFilterAction             permit
  IpFilterLogging            no
}
```

   Because IPSec is not required in this example, no filters for IKE traffic are needed.

8. Include the IpFilterRule statements in the IpFilterPolicy block.

   The IP filter rules and their relative placement within the IpFilterPolicy block should be from most specific to least specific. Because the AdminFTP rule controls traffic from a specific host, it should be placed before the InternalNetWeb rule. Note that to enable logging of the individual rules, filter logging must be enabled at the global level of the IP filter policy with the FilterLogging parameter.

```
IpFilterPolicy
{
   FilterLogging            on

   IpFilterRule             AdminFTP
   {
      IpSourceAddr          9.1.1.1
      IpDestAddr            9.1.1.2
      IpService
      {
         SourcePortRange       21
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpService
      {
         SourcePortRange       20
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional OutboundConnect
         Routing               local
         SecurityClass         1
      }
      IpGenericFilterActionRef  permit-log
   }
   IpFilterRule             InternalNetWeb
   {
      IpSourceAddr          9.1.1.1
      IpDestAddrSet         9.1.1.0/24
      IpService             WebServer
      {
         SourcePortRange       80
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpGenericFilterActionRef  permit-nolog
   }
}
```

9. Include all configured statements in the stack-specific IP security configuration file.

   The IpFilterPolicy statement and the IpGenericFilterAction statements are placed in the file in no particular order, although the file is easier to read if logically related items are placed close together. For further ease of reading and maintenance, document the file with comments, which begin with the number sign (#).

   The completed stack-specific IP security configuration file for the internal network with two filter rules follows:

```
# IP Security policy for Secure Server
##############################
# IpFilterPolicy block   #
##############################
IpFilterPolicy
{
   FilterLogging            on
   #Allow admin FTP; log traffic
   IpFilterRule             AdminFTP
   {
      IpSourceAddr          9.1.1.1
      IpDestAddr            9.1.1.2
      IpService
      {
         SourcePortRange       21
```

```
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpService
      {
         SourcePortRange       20
         DestinationPortRange  1024 65535
         Protocol              tcp
         Routing               local
         SecurityClass         1
         Direction             bidirectional OutboundConnect
      }
      IpGenericFilterActionRef  permit-log
   }
   #Allow LAN Web traffic; don't log
   IpFilterRule             InternalNetWeb
   {
      IpSourceAddr          9.1.1.1
      IpDestAddrSet         9.1.1.0/24
      IpService
      {
         SourcePortRange       80
         DestinationPortRange  1024 65535
         Protocol              tcp
         Direction             bidirectional InboundConnect
         Routing               local
         SecurityClass         1
      }
      IpGenericFilterActionRef  permit-nolog
}

##############################
# Generic Filter Actions   #
##############################
IpGenericFilterAction    permit-log
{
   IpFilterAction        permit
   IpFilterLogging       yes
}

IpGenericFilterAction    permit-nolog
{
   IpFilterAction        permit
   IpFilterLogging       no
}
```

10. Define an IP filter group for each zone, and include the IP filter rules that belong to that zone.

    In step “9” on page 979, both IpFilterRule statements include a reference to statements defined outside of the IpFilterRule block, the IpGenericFilterAction statements. Some other information, such as IP addresses and services, is undoubtedly to be needed more than once. Changing these occurrences to reference objects eliminates repeated typing of the same information and adds clarity to the configuration file. To take advantage of references, the reusable statements must be given a name.

    - Single IP addresses are defined by the IpAddr statement, which contains one parameter, Addr:

      ```
      IpAddr         InternalNetServerAddress
      {
         Addr        9.1.1.1
      {

      IpAddr         InternalNetAdminAddress
      {
         Addr        9.1.1.2
      }
      ```

    - Ranges or subnetworks are defined by the IpAddrSet statement, which contains either a Range or Prefix attribute:

      ```
      IpAddrSet      InternalNet
      {
         Prefix      9.1.1.0/24
      }
      ```

- To be referenced, each IpService statement needs a name:

```
IpService      WebServer
{
   SourcePortRange       80
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional InboundConnect
   Routing               local
   SecurityClass         1
}

IpService     FTPServer-Control
{
   SourcePortRange       21
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional InboundConnect
   Routing               local
   SecurityClass         1
}
IpService     FTPServer-Data
{
   SourcePortRange       20
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional OutboundConnect
   Routing               local
   SecurityClass         1
}
```

- FTP is composed of two individual services, and both can be condensed into an IpServiceGroup that references the two FTP services:

```
IpServiceGroup    FTPServer
{
   IpServiceRef   FTPServer-Control
   IpServiceRef   FTPServer-Data
}
```

- The IP filter rules can be grouped as well. Because the filter rules that apply to the internal network naturally relate to each other in the sense that they apply to the same security zone, they can be combined into an IpFilterGroup statement:

```
IpFilterGroup     InternalNetZoneA
{
   IpFilterRef    AdminFTP
   IpFilterRef    InternalNetWeb
}
```

  Notice that just as the list of IpFilterRule statements in the IpFilterPolicy block is ordered, the list of IpFilterRef statements in the IpFilterGroup block is also ordered. The InternalNetWeb rule applies to all of the IP addresses in the network, including the administrative machine. However, the AdminFTP rule is more specific because it applies only to a specific address within that network. The more specific rule is placed first in the list.

Now that all reusable statements have been identified and separately defined, they can be incorporated into any statement that requires that reusable statement type. The modified stack-specific IP security configuration file using references follows. Note that by adding names and organizing related statements, the purpose of the IpFilterPolicy statement is clarified.

```
# IP Security policy for Secure Server
##############################
# IpFilterPolicy block   #
##############################
IpFilterPolicy
{
   FilterLogging            on
   IpFilterGroupRef     InternalNetZoneA
}

##############################
# Security Zones         #
##############################
```

```
IpFilterGroup     InternalNetZoneA
{
   IpFilterRuleRef    AdminFTP
   IpFilterRuleRef    InternalNetWeb
}

############################
# Filter rules           #
############################
#Allow admin FTP; log traffic
IpFilterRule                 AdminFTP
{
   IpSourceAddrRef           InternalNetServerAddress
   IpDestAddrRef             InternalNetAdminAddress
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  permit-log
}

#Allow LAN Web traffic; don't log
IpFilterRule                 InternalNetWeb
{
   IpSourceAddrRef           InternalNetServerAddress
   IpDestAddrSetRef          InternalNet
   IpServiceRef              WebServer
   IpGenericFilterActionRef  permit-nolog
}

######## All reusable reference statements defined below ########

############################
# Generic Filter Actions   #
############################
IpGenericFilterAction    permit-log
{
   IpFilterAction        permit
   IpFilterLogging       yes
}

IpGenericFilterAction    permit-nolog
{
   IpFilterAction        permit
   IpFilterLogging       no
}


############################
# Reusable Services      #
############################
IpService      WebServer
{
   SourcePortRange       80
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional InboundConnect
   Routing               local
   SecurityClass         1
}

IpService     FTPServer-Control
{
   SourcePortRange       21
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional InboundConnect
   Routing               local
   SecurityClass         1
}

IpService     FTPServer-Data
{
   SourcePortRange       20
   DestinationPortRange  1024 65535
   Protocol              tcp
   Direction             bidirectional OutboundConnect
   Routing               local
   SecurityClass         1
}

############################
# Reusable Service Groups  #
############################
IpServiceGroup    FTPServer
```

```
{
   IpServiceRef   FTPServer-Control
   IpServiceRef   FTPServer-Data
}

################################
# Reusable IP Addresses    #
################################
IpAddr         InternalNetServerAddress
{
   Addr        9.1.1.1
}

IpAddr         InternalNetAdminAddress
{
   Addr        9.1.1.2
}

IpAddrSet      InternalNet
{
   Prefix      9.1.1.0/24
}
```

This stack-specific IP security configuration file gives FTP access to the administrator and web access to everyone in the internal network. By relying heavily on the abstracted use of references, the policy is not only more self-explanatory, but changes to any referenced object are propagated to any statement that references it. So, if the IP address of the administrative machine or internal subnetwork changes, you merely have to make one change to an IpAddr or IpAddrSet statement, rather than modify a large number of instances in multiple rules.

Using a common IP security configuration file for reusable statements

The stack-specific IP security configuration file should be tailored to the specific stack to which it belongs. However, as the policy files for IP security are being constructed, a large number of statements can be reused. Reusable statements can be placed in a common file, which is available to all stacks. Statements in the common IP security configuration file are read by all stacks on the system, providing a convenient way to store common definitions that they can all share. If you are operating in a sysplex, you can also place a common IP security configuration file on shared DASD or in a shared z/OS File System directory so that stacks in a multiple sysplex image have access to the same common configuration file.

Assuming that all statements that might be used later are placed in a common IP security configuration file, the stack-specific IP security configuration file from step 10 now reads as follows:

```
# IP Security policy for Secure Server
################################
# IpFilterPolicy block   #
################################
IpFilterPolicy
{
   FilterLogging            on
   IpFilterGroupRef     InternalNetZoneA
}

################################
# Security Zones         #
################################
IpFilterGroup     InternalNetZoneA
{
   IpFilterRuleRef    AdminFTP
   IpFilterRuleRef    InternalNetWeb
}

################################
# Filter rules           #
################################
#Allow admin FTP; log traffic
IpFilterRule                 AdminFTP
{
   IpSourceAddrRef           InternalNetServerAddress
   IpDestAddrRef             InternalNetAdminAddress
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  permit-log
}
```

```
#Allow LAN Web traffic; don't log
IpFilterRule                InternalNetWeb
{
   IpSourceAddrRef          InternalNetServerAddress
   IpDestAddrSetRef         InternalNet
   IpServiceRef             WebServer
   IpGenericFilterActionRef  permit-nolog
}
```

This stack-specific IP security configuration file references the following reusable statements:

- InternalNetServerAddress
- InternalNetAdminAddress
- InternalNet
- FTPServer
- FTPServer-Control
- FTPServer-Data
- permit-log
- permit-nolog
- WebServer

IpFilterRule statements can also be placed in the common IP security configuration file, because some IP filter rules apply to all addresses. If certain IpFilterRule statements are to apply globally to all stacks on the system, they can go into the common file. Use a value of `all` for the IpSourceAddr and IpDestAddr attributes. For instance, if all stacks on a z/OS system need rules permitting dynamic routing traffic (OSPF or RIP, for example), the statements that define this type of traffic can be placed in the common file and referenced in the stack-specific file:

```
IpFilterRule                    AllowOmprouteLocalNolog
{
  IpSourceAddr                  all
  IpDestAddr                    all
  IpServiceGroupRef             Omproute-local
  IpGenericFilterActionRef      Permit-nolog
}

IpServiceGroup                  Omproute-local
{
  IpServiceRef                  OSPF-local
  IpServiceRef                  RIP-local
}

IpService                       OSPF-local
{
  Protocol                      OSPF
  Direction                     bidirectional
  Routing                       local
}

IpService                       RIP-local
{
  Protocol                      UDP
  SourcePortRange               520
  DestinationPortRange          520
  Direction                     bidirectional
  Routing                       local
}
```

With these definitions in the common IP security configuration file, any stack needing global permission to send and receive routing information merely needs to include the following statement in the IpFilterPolicy block of its stack-specific IP security configuration file:

```
IpFilterRuleRef  AllowOmprouteLocalNolog
```

Steps for configuring the partner company model (host-to-host with IPSec)

The partner company model consists of two interconnected networks, with the server protecting traffic that originates from hosts outside the internal network.

Before you begin

The following statements and concepts are covered in the discussion of this model:

- Dynamic host-to-host IKE negotiations
- Key exchange rules
- Local and remote security endpoints
- Use of wildcards in Location and Identity
- IpLocalStartAction
- Granularity
- RSA signature peer authentication
- Certificates and certificate authorities
- On-demand activation
- Remote activation

Figure 125 on page 985 shows the partner company portion of the security model network.

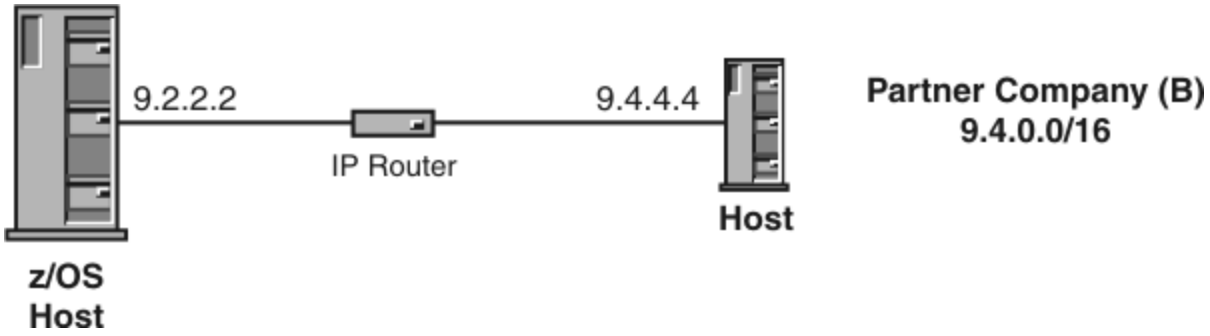


Figure 125. Partner company model

The partner company model is similar to that of the internal network, but an increased need for security means that a greater number of connections rely on data authentication and encryption. Although some services are allowed in the clear, sensitive data needs IPSec protection. To apply IPSec to a connection, the traffic that flows over that connection must match an IP filter rule that has an ipsec action.

For this example, assume you must meet the following requirements to allow network communications from a partner company in an untrusted zone B over a connected network (9.4.0.0/16) to a public IP address (9.2.2.2) on this host:

- Allow IKE traffic from untrusted zone B to this host.
- Allow secure FTP traffic (using TLS/SSL) from untrusted zone B to a secure FTP server running on this host.

  Secure FTP has its own security mechanism. Although IPSec can be used together with TLS/SSL, using both adds processing expense. For this example, secure FTP is allowed without IPSec protection.
- Allow Enterprise Extender (EE) traffic from untrusted zone B to an EE service running on this host using a dynamic IPSec tunnel with strong encryption and authentication.

  Because there is no encryption mechanism used in this example for EE, IPSec provides the secure service.
- Allow FTP traffic from untrusted zone B to an FTP server running on this host using a dynamic IPSec tunnel with strong AH authentication.

- The dynamic IPSec tunnel for EE comes up when outbound EE traffic is detected (on-demand activation).
- A dynamic IPSec tunnel for normal FTP control activates for each remote host that initiates an FTP connection (remote activation).
- A dynamic IPSec tunnel for normal FTP data activates for each outbound data connection to a remote host (local activation).
- Peers authenticate themselves using the RSA signature method.

Procedure

Perform the following steps to meet these requirements and configure the partner company model.

1. For each zone, determine what services are allowed:
   - IKE traffic
   - Normal FTP traffic
   - Secure FTP traffic
   - Enterprise Extender traffic
2. Define an IpService statement for each wanted service.
   - IKE uses UDP, port 500 for message exchanges:

```
IpService                   IKE-local
{
  SourcePortRange           500
  DestinationPortRange      500
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

   - For normal FTP traffic, allow inbound connections but do not allow outbound connection requests, other than data. Two services are required, one for the control connection and one for the data connection:

```
IpService                   FTPServer-Control
{
  SourcePortRange           21
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   FTPServer-Data
{
  SourcePortRange           20
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional OutboundConnect
  Routing                   local
  SecurityClass             0
}
```

   - For secure FTP traffic, allow inbound connections but do not allow outbound connection requests, other than data. Two services are required, one for the control connection and one for the data connection.

```
IpService                   SecureFTPServer-Control
{
  SourcePortRange           990
  DestinationPortRange      1024 65535
```

```
  Protocol                    tcp
  Direction                   bidirectional InboundConnect
  Routing                     local
  SecurityClass               0
}

IpService                     SecureFTPServer-Data
{
  SourcePortRange             989
  DestinationPortRange        1024 65535
  Protocol                    tcp
  Direction                   bidirectional OutboundConnect
  Routing                     local
  SecurityClass               0
}
```

- For Enterprise Extender traffic, allow both inbound and outbound traffic:

```
IpService                     Enterprise-Extender
{
  SourcePortRange             12000 12004
  DestinationPortRange        12000 12004
  Protocol                    UDP
  Direction                   bidirectional
  Routing                     local
  SecurityClass               0
}
```

IP services can be grouped together for convenience, flexibility, and reuse of configuration. Any IP filter rule that includes an IP service group will be expanded to include all of the services from that group. The following IP service groups include the IP services that define the FTPServer and the SecureFTPServer:

```
IpServiceGroup FTPServer
{
   IpServiceRef FTPServer-Control
   IpServiceRef FTPServer-Data
}
IpServiceGroup SecureFTPServer
{
   IpServiceRef Secure-FTPServer-Control
   IpServiceRef Secure-FTPServer-Data
}
```

3. Determine the data endpoints to be protected.

   In this example, the local data endpoint is the public address of the server, and the remote data endpoint is the entire subnetwork in zone B:

```
Local public IP address:  9.2.2.2
Remote subnet: 9.4.0.0/16
```

4. Determine what level of security is needed between each set of data endpoints:

```
IKE traffic - permit
Secure FTP traffic - permit
EE traffic - IPSec encryption and authentication
FTP traffic - IPSec authentication
```

5. Configure an IpGenericFilterAction statement for the level of security that is required.

   The requirements stated that IKE and secure FTP should be allowed, and that EE and normal FTP traffic required IPSec protection. Notice that the parameters of the IpGenericFilterAction statement include ipsec as well as permit and deny. Because no logging requirements were specified, two actions are needed, permit and ipsec:

```
IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
```

```
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
   IpFilterLogging       no
}
```

6. If IPSec is required between any two endpoints, take the following actions:
   a) Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:
      i) Determine the required type and strength of protection for the phase 1 Security Association. Phase 1 Security Associations must use both authentication and encryption. Because strong encryption was specified for phase 2, use strong encryption for phase 1:

```
Authentication, SHA1 algorithm
Encryption, 3DES algorithm
DHGroup, Diffie-Hellman Group2
```

      Tip: The greater the Diffie-Hellman group specified, the greater the protection provided.

      ii) Decide what type of peer authentication is used:

```
RSA signature
```

      Because the remote data endpoint represents multiple hosts in this example, RSA signature is used. RSA signature authentication provides greater flexibility, scalability, and security than pre-shared key authentication. Because there are potentially multiple remote peers in the partner company's subnet, RSA signature is a reasonable choice.

      The use of RSA signature requires setup of RACF certificates and certificate authority information. Both IKE peers need access to a key ring with at least one X.509 digital certificate to identify itself. To set up the IKE daemon for certificates, see Appendix E, “Steps for preparing to run IP security,” on page 1361. The site should decide what certificate authorities are recognized. A certificate authority can be an outside commercial entity, or it can be defined locally in RACF. For information about installing certificate authorities in RACF, see z/OS Security Server RACF Security Administrator's Guide.

      In this example, the IKEv1 protocol is used and the IKE daemon is using the native certificate service. A certificate authority with label CA4PartnerCompany is used, which is presumed to have been defined as a certificate authority in the RACF database. The label CA4PartnerCompany should be added to the iked.conf file as a recognized certificate authority as follows:

```
SupportedCertAuth CA4PartnerCompany
```

      Guideline: For more efficient processing of certificates when the IKED is using the native certificate service, you should code SupportedCertAuth for each acceptable certificate authority that will be used to sign certificates for remote IKE peers.

      iii) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation, including type of peer authentication, strength of encryption, and how often the phase 1 keys are refreshed. Because KeyExchangeOffer statements are reusable, give it a descriptive name as follows:

```
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RsaSignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
```

```
   RefreshLifesizeAccepted  none
}
```

iv) Decide whether NAT traversal will be allowed.

In this example, the following parameter is used:

```
AllowNat No
```

v) Determine the negotiation mode. Main mode is used for phase 1, providing added security by encrypting the identities of the two IKE peers during the phase 1 negotiation.

vi) Configure a KeyExchangeAction statement that defines the control information for the phase 1 negotiation. The key exchange action determines the mode of the phase 1 negotiation and the parameters that are included in the KeyExchangeOffer statement. Although multiple KeyExchangeOffer statements are acceptable, only one is required. Both peers must agree on the parameters. Use the KeyExchangeOffer statement that was configured in step "7.a.iii" on page 988:

```
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate       main
   HowToRespondIKEv1   main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}
```

vii) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement. In this example, the local security endpoint is the local host. An identity and IP address of the local IKE peer is required as follows:

```
LocalSecurityEndpoint  Public_IKED
{
   Identity      IpAddr  9.2.2.2
   Location      9.2.2.2
}
```

Similarly, the same information is needed for remote hosts. Notice that in the RemoteSecurityEndpoint statement, an IP address range is specified for both identity and location. Although it is possible to configure a RemoteSecurityEndpoint statement for every host in the remote subnetwork, unless they have unique key exchange requirements, it is not necessary. Wildcards can be used for the Identity and Location parameters. Any of the identity types can potentially contain wildcards to include a group of remote hosts with similar identities.

```
RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      IpAddr   9.4.0.0/16
   Location      9.4.0.0/16
   CaLabel       CA4PartnerCompany
}
```

The use of an IPv4 address on the Identity parameter for the Local_IKED security endpoint requires that the X.509 digital certificate for the local IKE daemon include the IPv4 address in the Subject Alternative Name field of the certificate. The use of an IPv4 subnet on the Identity parameter for the ZoneB_IKED security endpoint requires that an IPv4 address within that subnet (9.4.0.0/16) appear in the Subject Alternative Name field of the X.509 digital certificate for the remote IKE daemon.

The inclusion of the CaLabel parameter in the ZoneB_IKED security endpoint emphasizes the fact that the local host requests that the remote host use only certificates that are signed by the certificate authority that is identified by the label CA4PartnerCompany.

Rule: For RSA signature mode authentication, the identity of a security endpoint must be contained in its X.509 digital certificate, either in the Subject Name field or the Subject Alternative Name field.

For detailed specifications on the use of wildcards in the RemoteSecurityEndpoint statement, see z/OS Communications Server: IP Configuration Reference.

viii) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action:

```
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}
```

ix) Include the key exchange rule in the KeyExchangePolicy statement:

```
KeyExchangePolicy
{
   KeyExchangeRuleRef    ZoneB_KeyExRule1
}
```

b) Configure an IpDynVpnAction statement that defines the parameters of the phase 2 negotiation as follows:

i) Determine the required type and strength of IPSec protection for the phase 2 Security Association.

In this example, there is a unique requirement for each traffic type. Therefore, there are two types of IPSec protection for each type of traffic as follows:

```
EE traffic - strong encryption: ESP, 3DES
             strong ESP authentication: Hmac SHA1
Normal FTP traffic - strong AH authentication, Hmac SHA1
```

This information eventually translates into two IpDataOffer statements.

ii) Determine whether tunnel or transport mode is required.

Tunnel mode is required only if either endpoint of the Security Association is a secure gateway. The choice is optional in a host-to-host configuration, but transport mode is typically used.

iii) Configure IpDataOffer statements that define the parameters of the phase 2 negotiation.

- Authenticated offer for FTP:

```
IpDataOffer TRAN-AHSHA-NOENCR
{
   HowToEncap Transport
   HowToEncrypt DoNot
   HowToAuth    AH HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

- Encrypted and authenticated offer for EE:

```
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    ESP HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
```

```
   RefreshLifesizeAccepted  none
}
```

iv) Determine which peer is allowed to initiate the negotiation.

Because the EE VPN activates when outbound EE traffic is detected, you need to be able to start the negotiation locally. The remote host initiates the negotiation of the Security Association for the FTP control connection, so local initiation is not required. The FTP data connection is initiated from the local host, so to activate the Security Association, local initiation is required.

v) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation.

The IpDynVpnAction statement can control which host is allowed to initiate the negotiation. You must allow the IPSec VPN for the FTP control connection to be initiated by the remote peer, and the data connection by the local host. EE should be allowed to initiate an IKE negotiation locally.

- Authenticated VPN action for FTP:

```
IpDynVpnAction FTP-vpnaction
{
   Initiation           either
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-AHSHA-NOENCR
}
```

- Encrypted and authenticated VPN action for EE traffic:

```
IpDynVpnAction EE-vpnaction
{
   Initiation           localonly
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-ESPSHA-3DES
}
```

c) Decide how the Security Association is to be activated as follows:

i) Configure an optional LocalDynVpnPolicy statement, if command-line activated or autoactivated.

Neither the EE nor FTP VPNs require a LocalDynVpnPolicy statement, because neither is command-line activated or autoactivated.

ii) Configure an optional IpLocalStartAction statement, if the Security Association is to be activated locally (that is, on-demand, command-line, or autoactivation) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host Security Association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

Before a phase 2 negotiation can initiate, the IKE daemon needs to know the IP addresses, ports, and protocols that the Security Association covers. In most cases, these can be inferred from the filter rule, or from the packet that started the on-demand activation. An IpLocalStartAction statement explicitly defines where these parameters are obtained. The granularity setting determines whether the information comes from the matching filter rule, or from the packet. By explicitly specifying packet, you can guarantee that a new Security Association is created for each connection request.

```
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
```

```
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranularity       packet
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}
```

In this example, specifying `packet` is not required. IKE detects that the filter rule has a range of ports specified and resorts to packet granularity instead of the default of `rule`. Specifying `packet` is required, however, if the EE and FTP rules had all ports specified. In that case, one Security Association is negotiated for all ports, not a single port. To avoid confusion, it is better to specify your intention.

iii) Set the global PreDecap parameter of the IpFilterPolicy statement to `off`, or create an IpFilterRule statement that allows IPSec traffic (AH and ESP).

When the system begins to process encapsulated traffic, the encapsulated packets are subject to filtering. The encrypted and authenticated packets are denied unless they are explicitly allowed. There are two options to configure this that are subtly different. The first is less restrictive, allowing AH and ESP packets from anywhere and not subjecting them to filtering until after the packets have had their IPSec headers removed. The second method is to add an IpFilterRule statement that explicitly permits the IPSec-encapsulated traffic. This option adds an additional layer of security, in that you can control exactly who is allowed to send encrypted traffic. This protects system resources by preventing the processing of IPSec packets unnecessarily, such as might happen in the case of a malicious attack of IPSec packets. Although more secure, this solution is less efficient because it requires additional processing resources to filter all IPSec-encapsulated traffic.

For the first option, in the main IpFilterPolicy block, code `PreDecap off` as follows:

```
IpFilterPolicy
{
   PreDecap off
   ⋮
}
```

For the second option, configure a filter rule in the IpFilterPolicy block that explicitly allows AH and ESP traffic as follows:

```
IpFilterRule           Allow-IPSec-traffic
{
   IpSourceAddr        9.2.2.2
   IpDestAddrSet       9.4.0.0/16
   IpService                   AH-traffic
   {
      Protocol                 AH
      Direction                bidirectional
      Routing                  local
      SecurityClass            0
   }
   IpService                   ESP-traffic
   {
      Protocol                 ESP
      Direction                bidirectional
      Routing                  local
      SecurityClass            0
   }
   IpGenericFilterRef    permit-nolog
}
```

Tip: In most cases, unless the extra security is deemed a necessity, use the `PreDecap off` global option to allow IPSec packets to flow with minimum overhead.

7. Define an IpFilterRule statement for each set of data endpoints.

The secure host and subnetwork B represent the data endpoints. The IpService statements were defined in step “2” on page 986. Place the IpDynVpnAction statements that were created in step “6” on page 988 with the appropriate rule.

```
IpFilterRule              ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}


IpFilterRule              FTPServer-ZoneB
{
   IpSourceAddr              9.2.2.2
   IpDestAddrSet             9.4.0.0/16
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            FTP-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}
IpFilterRule             EE-ZoneB
{
   IpSourceAddr          9.2.2.2
   IpDestAddrSet         9.4.0.0/16
   IpService
   {
      SourcePortRange       12000 12004
      DestinationPortRange  12000 12004
      Protocol              udp
      Direction             bidirectional
      Routing               local
      SecurityClass         0
   }
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}
```

Because both IKE and Secure FTP need to be permitted without IPSec protection, an IpFilterRule statement for both is not needed. The two services can be combined into one rule.

Tip: Any services that share the same data endpoints and the same security requirements can be placed together in one IpFilterRule statement.

8. Include the IpFilterRule statements in the IpFilterPolicy block.

   The IpFilterRule statement allowing IKE traffic should always be at the top of the list. The rest of the IpFilterRule statements are disjointed, and their relative placement within the IpFilterPolicy is irrelevant.

   Tip: If it is known that one particular type of traffic is the most frequent, placing that rule near the top of the list results in faster filter lookups.

9. Define an IP filter group for each zone and include the IpFilterRule statements that belong to that zone.

   Although the creation of reference objects and groups is not mandatory, they provide for ease of maintenance as the IP security policy grows more complex.

   A completely configured policy, including all objects and their references, is as follows:

```
# IpFilterPolicy for secure public server

IpFilterPolicy
{
   PreDecap             off
   IpFilterGroupRef     ZoneB
}
```

```
KeyExchangePolicy
{
   KeyExchangeRuleRef   ZoneB_KeyExRule1
}

####### All reusable statements follow ########
IpFilterGroup           ZoneB
{
   IpFilterRuleRef      ZoneB-Permitted-traffic
   IpFilterRuleRef      FTPServer-ZoneB  #IPSec-protected
   IpFilterRuleRef      EE-ZoneB         #IPSec-protected
}

##########################################
# IpFilterRules                          #
#   defines:                             #
#      data endpoints                    #
#      Allowed services                  #
#      Actions (permit, deny, ipsec)     #
##########################################
IpFilterRule            ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}

IpFilterRule           EE-ZoneB
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              Enterprise-Extender
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}

IpFilterRule           FTPServer-ZoneB
{
   IpSourceAddrRef           PublicServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         FTP-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}

############################
# Local Start Actions #
############################
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranularity       packet
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}

########################
# IpService groups #
########################
IpServiceGroup            FTPServer
{
   IpServiceRef           FTPServer-Control
   IpServiceRef           FTPServer-Data
```

```
}

IpServiceGroup            SecureFTPServer
{
   IpServiceRef           SecureFTPServer-Control
   IpServiceRef           SecureFTPServer-Data
}

#######################################
# Services provided by this host #
#######################################

IpService                    IKE-local
{
  SourcePortRange            500
  DestinationPortRange       500
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    SecureFTPServer-Control
{
  SourcePortRange            990
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    SecureFTPServer-Data
{
  SourcePortRange            989
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional OutboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    Enterprise-Extender
{
  SourcePortRange            12000 12004
  DestinationPortRange       12000 12004
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    FTPServer-Control
{
  SourcePortRange            21
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    FTPServer-Data
{
  SourcePortRange            20
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional OutboundConnect
  Routing                    local
  SecurityClass              0
}
```

```
##########################
# Security Endpoints #
##########################
LocalSecurityEndpoint  Public_IKED
{
   Identity      IpAddr  9.2.2.2
   Location      9.2.2.2
}

RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      IpAddr   9.4.0.0/16
   Location      9.4.0.0/16
   CaLabel       CA4PartnerCompany
}

#############################
# Generic filter actions #
#############################

IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
   IpFilterLogging       no
}

#####################################
# Key Exchange offers            #
#   defines:                     #
#     Authentication type        #
#     Encryption type            #
#     Peer authentication method #
#     Refresh limits             #
#####################################
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RsaSignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

#####################################
# Key Exchange Actions           #
#  defines:                      #
#    Negotiation mode            #
#    List of Key exchange offers #
#####################################
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate       main
   HowToRespondIKEv1   main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}

#########################################
# KeyExchangeRules                   #
#   defines:                         #
#      A pair of security endpoints  #
#      permitted in IKE negotiations #
#########################################
KeyExchangeRule              ZoneB_KeyExRule1
```

```
{
   LocalSecurityEndpointRef  Public_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}

##################################
# Data Offers                    #
#   defines:                     #
#      Encapsulation mode        #
#      Authentication type       #
#      Encryption type           #
#      Refresh limits            #
##################################
### Authenticated offer ###
IpDataOffer TRAN-AHSHA-NOENCR
{
   HowToEncap Transport
   HowToEncrypt DoNot
   HowToAuth    AH HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

### Encrypted offer ###
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    DoNot
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

#####################################
# Dynamic VPN Actions               #
#   defines:                        #
#     Initiation role               #
#     Pfs group                     #
#     Lifetime of connection        #
#     List of Data offers           #
#####################################
IpDynVpnAction FTP-vpnaction
{
   Initiation           either
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-AHSHA-NOENCR
}

IpDynVpnAction EE-vpnaction
{
   Initiation           localonly
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-ESPSHA-3DES
}

###################
# IP addresses #
###################

IpAddr    PublicServerAddress
{
   Addr   9.2.2.2
}
```

```
IpAddrSet ZoneB-subnet
{
   Prefix 9.4.0.0/16
}
```

10. Include all configured statements in the stack-specific IP security configuration file.

Steps for configuring the partner company with NAT model (host-to-host with IPSec)

In the partner company with NAT model, the partner company model topology is modified to include private addressing in the private network of each partner company, with a NAT device in front of each private network.

Before you begin

The following statements and concepts are covered in the discussion of this model:

- AllowNat and NatKeepAliveInterval parameters on the KeyExchangePolicy and KeyExchangeAction statements
- IKE traffic on UDP port 4500, in addition to port 500
- NAT implications for host-to-host dynamic IKE negotiations:
  - Local and remote data endpoints
  - Local and remote security endpoints
  - IKE initiator and responder roles
  - Restriction on HowToAuth protocol (AH not supported)
- Using wildcards for location and identity
- RSA signature peer authentication
- Certificates and certificate authorities
- CaLabel
- SupportedCertAuth

"Steps for configuring the partner company model (host-to-host with IPSec)" on page 985 assumed a network topology with both partner companies using public IP addresses in their internal networks. Often one or both businesses have an internal network that uses IETF-defined private IP addresses (10.0.0.0/8, 172.16.0.0./12, and 192.168.0.0/16). Private IP addresses cannot be routed outside an internal network. Network address translation (NAT) is used to create a mapping of private addresses to public addresses and perform the necessary translation as packets traverse the NAT device.

When IPSec Security Associations traverse a NAT, there are problems because the NAT is unable to update IP addresses and checksums that are part of the encapsulated data (encrypted, authenticated, or both). The IETF has defined a solution known as NAT traversal (NATT) that allows IPSec Security Associations to successfully traverse a NAT device.

Figure 126 on page 999 shows the partner company with NAT topology when the partner company model topology has been modified to include private addressing in each partner company's private network with a NAT device in front of each private network.

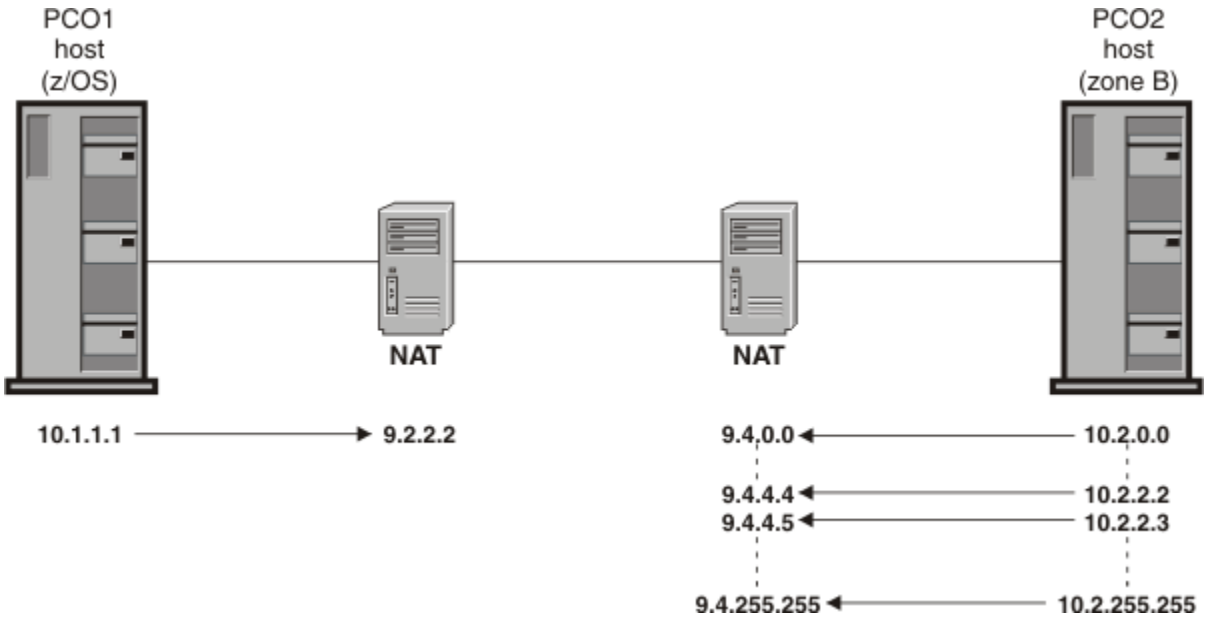


Figure 126. Partner company with NAT model

The steps in this topic will describe the configuration considerations and requirements when the NATT solution is implemented to traverse NAT devices in a host-to-host environment. The partner company with NAT model has the same basic security requirements as the partner company model. Configuration statements added or changed for the partner company with NAT model are shown in **bold**. The example describes the policy for partner company 1 (PCO1).

For this example, assume you must meet the following requirements to allow network communications from a partner company (PCO2) in an untrusted zone B behind a NAT over a connected network (9.4.0.0/16) to a server on this host that is behind a NAT:

- IKE traffic from untrusted zone B that is behind a NAT is allowed to this host that is behind a NAT.
- Secure FTP traffic (using TLS/SSL) from untrusted zone B is allowed to a secure FTP server running on this host.
- Enterprise Extender (EE) traffic from untrusted zone B is allowed to an EE service running on this host using a dynamic IPSec tunnel with strong encryption and authentication.
- FTP traffic from untrusted zone B is allowed to an FTP server running on this host using a dynamic IPSec tunnel with strong authentication.
- The dynamic IPSec tunnel for EE is activated when outbound EE traffic is detected (on-demand activation).
- A dynamic IPSec tunnel for normal FTP control activates for each remote host that initiates an FTP connection (remote activation).
- A dynamic IPSec tunnel for normal FTP data activates for each remote host that initiates an FTP data connection (remote activation).
- Peers authenticate themselves using the RSA signature method.

Procedure

Perform the following steps to meet these requirements and configure the partner company with NAT model.

1. For each zone, determine what services are allowed:
   - IKE traffic
   - Normal FTP traffic
   - Secure FTP traffic

- Enterprise Extender traffic

2. Define an IpService statement for each wanted service.
   - IKE traffic

     When NAT traversal is allowed, IKE uses UDP port 500 and port 4500 for message exchanges. NAT traversal is controlled by the AllowNat parameter on the KeyExchangePolicy and KeyExchangeAction statements.

```
IpService                   IKE-local-500
{
  SourcePortRange           500
  DestinationPortRange      500
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}

IpService                   IKE-local-4500
{
  SourcePortRange           4500
  DestinationPortRange      4500
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

   - Normal FTP traffic

     When a NAT is being traversed, FTP clients typically need to use passive mode (PASV) or extended passive mode (EPSV) to connect to the FTP server, allowing inbound connections but not outbound connection requests. For more information on active and passive mode FTP, see "Considerations for IPSec-encapsulated FTP traffic when traversing a NAT" on page 1039.

     Two services are required, one for the control connection and one for the data connection. The range of server ports specified for the data connection reflects the port range specified for PASSIVEDATAPORTS in the server's FTP.DATA file. For more information on PASSIVEDATAPORTS, see z/OS Communications Server: IP Configuration Reference.

     The following definitions show the services required for the FTP server for EPSV mode:

```
IpService                   FTPServer-Control
{
  SourcePortRange           21
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   FTPServer-Data-Passive
{
  SourcePortRange           50000 50200
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}
```

   - Secure FTP traffic

     Again, when a NAT is being traversed, FTP clients typically need to use passive mode (PASV) or extended passive mode (EPSV) to connect to the FTP server, allowing inbound connections but not

outbound connection requests. Two services are required, one for the control connection and one for the data connection.

```
IpService                   SecureFTPServer-Control
{
  SourcePortRange           990
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}

IpService                   SecureFTPServer-Data-Passive
{
  SourcePortRange           50201 50400
  DestinationPortRange      1024 65535
  Protocol                  tcp
  Direction                 bidirectional InboundConnect
  Routing                   local
  SecurityClass             0
}
```

- Enterprise Extender traffic

  Allow both inbound and outbound connections.

```
IpService                   Enterprise-Extender
{
  SourcePortRange           12000 12004
  DestinationPortRange      12000 12004
  Protocol                  UDP
  Direction                 bidirectional
  Routing                   local
  SecurityClass             0
}
```

3. Determine the data endpoints that are to be protected.

   In this case, the local data endpoint is the private address of the server, local IP address 10.1.1.1.

   The remote data endpoint is the partner company's internal network. The z/OS implementation does not require you to code private addresses for the remote endpoint; the network address translated public addresses should be configured as the remote data endpoint. In this example, the private addresses in the PCO2 internal network (10.2.0.0/16) are translated into the public address range 9.4.0.0/16. Thus, the remote subnetwork is 9.4.0.0/16.

4. Determine what level of security is needed between each set of data endpoints.

```
IKE traffic - permit
Secure FTP traffic - permit
EE traffic - IPSec encryption and authentication
FTP traffic - IPSec authentication
```

5. Configure an IpGenericFilterAction statement for the level of security that is required.

   The requirements stated that IKE and secure FTP should be allowed, and that EE and normal FTP traffic required IPSec protection. Because no logging requirements were specified, two actions are needed, permit and ipsec.

```
IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
```

```
   IpFilterLogging       no
}
```

6. If IPSec is required between any two endpoints, take the following actions:
   a) Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:
      i) Determine the required type and strength of protection for the phase 1 Security Association. Phase 1 Security Associations must use both authentication and encryption. Because strong encryption was specified for phase 2, use strong encryption for phase 1:

```
Authentication, SHA1 algorithm
Encryption, 3DES algorithm
DHGroup, Diffie-Hellman Group2
```

      Tip: The greater the Diffie-Hellman group specified, the greater the protection provided.

      ii) Decide what type of peer authentication is used:

```
RSA signature
```

      Because the remote data endpoint represents multiple hosts in this example, RSA signature is used. RSA signature authentication provides greater flexibility, scalability, and security than pre-shared key authentication. Because there are potentially multiple remote peers in the partner company's subnet, RSA signature is a reasonable choice.

      The use of RSA signature requires setup of RACF certificates and certificate authority information. Both IKE peers need access to a key ring with at least one X.509 digital certificate to identify itself. To set up the IKE daemon for certificates, see Appendix E, “Steps for preparing to run IP security,” on page 1361. The site should decide what certificate authorities are recognized. A certificate authority can be an outside commercial entity, or it can be defined locally in RACF. For information about installing certificate authorities in RACF, see z/OS Security Server RACF Security Administrator's Guide.

      In this example, the IKEv1 protocol is used and the IKE daemon is using the native certificate service. A certificate authority with label CA4PartnerCompany is used, which is presumed to have been defined as a certificate authority in the RACF database. The label CA4PartnerCompany should be added to the iked.conf file as a recognized certificate authority as follows:

```
SupportedCertAuth CA4PartnerCompany
```

      Guideline: For more efficient processing of certificates when the IKED is using the native certificate service, you should code SupportedCertAuth for each acceptable certificate authority that will be used to sign certificates for remote IKE peers.

      iii) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation, including type of peer authentication, strength of encryption, and how often the phase 1 keys are refreshed. Because KeyExchangeOffer statements are reusable, give it a descriptive name as follows:

```
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RSAsignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

      iv) Decide whether NAT traversal will be allowed.

The AllowNat parameter on the KeyExchangePolicy and KeyExchangeAction statements allows IKE to advertise NAT traversal support. When a phase 1 Security Association is being negotiated, if both IKE peers support NAT traversal, several IKE payloads are exchanged that allow the peers to determine if one or more NAT devices are being traversed. If a NAT is being traversed, the IKE peers negotiate a UDP-encapsulated transport mode or UDP-encapsulated tunnel mode phase 2 Security Association to allow IPSec traffic to traverse the NATs successfully. If a NAT is not being traversed, a standard transport or tunnel mode phase 2 Security Association is negotiated.

The AllowNat parameter can be specified on the overall KeyExchangePolicy statement or for a specific KeyExchangeAction statement. If AllowNat is No, IKE does not advertise its support for NAT traversal. You might want to disable NAT traversal for interoperability purposes (certain remote resources might not be able to tolerate NAT traversal protocols) or security concerns (for example, NAT traversal exposes private internal addresses to the IKE peer). If NAT traversal is disabled and there is a NAT device in the path, ipsec processing might fail to negotiate the Security Association or fail to send data over the Security Association.

For this model, allow NAT traversal support to be advertised for all phase 1 Security Associations by coding the following parameter on the KeyExchangePolicy statement:

```
AllowNat Yes
```

A NatKeepAliveInterval parameter is also provided on the KeyExchangePolicy statement. A NAT keep-alive timer is maintained to ensure that NAT mappings do not expire. If z/OS is behind a NAT, a NAT keep-alive timer is started with the interval specified on the NatKeepAliveInterval parameter. If z/OS is behind a NAT that is using static mappings that will not expire, the NatKeepAliveInterval parameter should be set to 0. It is not necessary to run a NAT keep-alive timer in this case.

For this model, because static mapping is being used for the NAT in front of PCO1, the NatKeepAliveInterval parameter is set to 0:

```
NatKeepAliveInterval   0
```

v) Determine the negotiation mode. Main mode is used for phase 1 in this example, providing added security by encrypting the identities of the two IKE peers during the phase 1 negotiation.

vi) Configure a KeyExchangeAction statement that defines the control information for the phase 1 negotiation. The key exchange action determines the mode of the phase 1 negotiation and the parameters that are included in the KeyExchangeOffer statement. Although multiple KeyExchangeOffer statements are acceptable, only one is required. Both peers must agree on the parameters. Use the KeyExchangeOffer statement that was configured previously:

```
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate       main
   HowToRespondIKEv1   main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}
```

vii) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement. In this example, the local security endpoint is the local host. An identity and IP address of the local IKE peer is required as follows:

```
LocalSecurityEndpoint  Local_IKED
{
   Identity      Fqdn  Server.PCO1.example.com
   Location      10.1.1.1
}
```

The value specified for Identity on the LocalSecurityEndpoint statement is sent to the remote peer to identify this endpoint. The remote peer uses this value to determine if a Security

Association can be negotiated and which policy should be used. Identity can be specified in different formats, such as an IP address or fully qualified domain name.

The partner company model specified an IP address for Identity, and an IP address could be used in this model. However, PCO1's private IP address 10.1.1.1 is not meaningful in the PCO2 configuration, so one of the other Identity types was chosen. The fully qualified domain name (Fqdn) identifies the local security endpoint.

An identity and IP address are also needed for the remote hosts. Notice that in the following RemoteSecurityEndpoint statement, the Identity and Location attributes use a wildcard to include a group of remote hosts with similar identities. It is possible to use a wildcard for the values because the key exchange requirements are the same for the hosts included in the specified range.

The Identity attribute is specified as Fqdn (fully qualified domain name), and the Location attribute is the range of public IP addresses used by the PCO2 internal network. PCO2 internal private addresses are not coded because they have no meaning for PCO1.

```
RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      Fqdn   *.PCO2.example.com
   Location      9.4.0.0/16
   CaLabel       CA4PartnerCompany
}
```

The use of a fully qualified domain name on the Identity parameter for the Local_IKED security endpoint requires that the X.509 digital certificate for the local IKE daemon include the fully-qualified domain name in the Subject Alternative Name field of the certificate. The use of a wild-carded domain name on the Identity parameter for the ZoneB_IKED security endpoint requires that a fully-qualified domain name ending with .PCO2.example.com appear in the Subject Alternative Name field of the X.509 digital certificate for the remote IKE daemon.

The inclusion of the CaLabel parameter in the ZoneB_IKED security endpoint emphasizes the fact that the local host requests that the remote host use only certificates that are signed by the certificate authority that is identified by the label CA4PartnerCompany.

Rule: For RSA signature mode authentication, the identity of a security endpoint must be contained in its X.509 digital certificate, either in the Subject Name field or the Subject Alternative Name field.

For detailed specifications on the use of wildcards in the RemoteSecurityEndpoint statement, see z/OS Communications Server: IP Configuration Reference.

viii) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action:

```
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}
```

ix) Include the key exchange rule in the KeyExchangePolicy statement block. Also include the AllowNat parameter:

```
KeyExchangePolicy
{
   AllowNat               Yes
   KeyExchangeRuleRef     ZoneB_KeyExRule1
}
```

b) Configure an IpDynVpnAction statement that defines the parameters of the phase 2 negotiation as follows:

i) Determine the required type and strength of IPSec protection for the phase 2 Security Association.

In this example, there is a unique requirement for each traffic type. Therefore, there are two types of IPSec protection for each type of traffic as follows:

```
EE traffic - strong encryption: ESP, 3DES
             strong ESP authentication: Hmac SHA1
Normal FTP traffic - strong ESP authentication, Hmac SHA1
```

The partner company model used AH authentication for normal FTP traffic, but the NATT solution defined by the IETF is based on the ESP protocol. The AH protocol is not supported for NATT, and ESP authentication is used in this model.

This information eventually translates into two IpDataOffer statements.

ii) Determine whether tunnel or transport mode is required.

Tunnel mode is required only if either endpoint of the Security Association is a secure gateway. The choice is optional in a host-to-host configuration, but transport mode is typically used.

iii) Configure IpDataOffer statements that define the parameters of the phase 2 negotiation.

- Authenticated offer for FTP:

```
IpDataOffer TRAN-ESPSHA-NOENCR
{
   HowToEncap Transport
   HowToEncrypt DoNot
   HowToAuth    ESP HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

- Encrypted and authenticated offer for EE:

```
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    ESP HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}
```

iv) Determine which peer is allowed to initiate the negotiation.

When an IKE peer is behind a NAT device, there are implications regarding which peer can initiate. In a host-to-host configuration, IKE should initiate only to a peer whose IP address is unambiguous. If the peer is not behind a NAT, or if the peer's NAT mapping is static, the address is unambiguous. In this model, assume that static mapping is being used for both PCO1 and PCO2.

Because the EE VPN activates when outbound EE traffic is detected, you need to be able to start the negotiation locally. For PASV mode or EPSV mode FTP, both the FTP control connection and the FTP data connection initiate the negotiation of the Security Association remotely, so local initiation is not required.

v) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation.

The IpDynVpnAction statement can control which host is allowed to initiate the negotiation. Both the FTP control connection and data connection will be initiated remotely. EE should be allowed to initiate an IKE negotiation locally.

- Authenticated VPN action for FTP:

```
IpDynVpnAction FTP-vpnaction
{
   Initiation            remoteonly
   InitiateWithPfs       group2
   AcceptablePfs         group2
   VpnLife               1440
   IpDataOfferRef        TRAN-ESPSHA-NOENCR
}
```

- Encrypted and authenticated VPN action for EE traffic:

```
IpDynVpnAction EE-vpnaction
{
   Initiation            localonly
   InitiateWithPfs       group2
   AcceptablePfs         group2
   VpnLife               1440
   IpDataOfferRef        TRAN-ESPSHA-3DES
}
```

c) Decide how the Security Association is to be activated as follows:

i) Configure an optional LocalDynVpnPolicy statement, if command-line activated or autoactivated.

Neither the EE nor FTP VPNs require a LocalDynVpnPolicy statement, because neither is command-line activated or autoactivated.

ii) Configure an optional IpLocalStartAction statement, if the Security Association is to be activated locally (that is, on-demand, command-line, or autoactivated) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host Security Association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

Before a phase 2 negotiation can initiate, the IKE daemon needs to know the IP addresses, ports, and protocols that the Security Association covers. In most cases, these can be inferred from the filter rule, or from the packet that started the on-demand activation. An IpLocalStartAction statement explicitly defines where these parameters are obtained. The granularity setting determines whether the information comes from the matching filter rule, or from the packet. By explicitly specifying packet, you can guarantee that a new Security Association is created for each connection request.

```
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranulatiry       packet
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}
```

In this example, specifying `packet` for LocalPortGranularity and RemotePortGranularity is not required. IKE detects that the filter rule has a range of ports specified and resorts to packet granularity instead of the default of `rule`. Specifying `packet` is required, however, if the EE rules had all ports specified. In that case, one Security Association is negotiated for all ports, not a single port. To avoid confusion, it is better to specify your intention.

iii) Set the global PreDecap parameter of the IpFilterPolicy statement to `off`, or create an IpFilterRule statement that allows IPSec traffic (AH and ESP).

In this example, set `PreDecap off` in the IpFilterPolicy statement

7. Define an IpFilterRule statement for each set of data endpoints.

The secure host and zone B represent the data endpoints. The IpService statements were defined previously. Place the IpDynVpnAction statements that were created previously with the appropriate rule. Notice that the parameters of the IpGenericFilterAction statement include ipsec as well as permit and deny. Because no logging requirements were specified, two actions are needed, permit and ipsec.

```
IpFilterRule              ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local-500
   IpServiceRef              IKE-local-4500
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}


IpFilterRule                 FTPServer-ZoneB
{
   IpSourceAddr              10.1.1.1
   IPDestAddrSet             9.4.0.0/16
   IpServiceRef              FTPServer-Control
   IpServiceRef              FTPServer-Data-Passive
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            FTP-vpnaction
}

IpFilterRule             EE-ZoneB
{
   IpSourceAddr              10.1.1.1
   IPDestAddrSet             9.4.0.0/16
   IpServiceRef              Enterprise-Extender
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnAction            EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}
```

Because IKE port 500, IKE port 4500, and secure FTP without IPSec protection are permitted, a separate IpFilterRule statement for each is not needed. The services can be combined into one IpFilterRule statement.

Tip: Any services that share the same data endpoints and the same security requirements can be placed together in one IpFilterRule statement.

8. Include the IpFilterRule statements in the IpFilterPolicy block.
9. Define an IP filter group for each zone and include the IpFilterRule statements that belong to that zone. Although the creation of reference objects and groups is not mandatory, they provide for ease of maintenance as the IP security policy grows more complex.
10. Include all configured statements in the stack-specific IP security configuration file.

Results

A completely configured policy, including all objects and their references, is as follows:

```
# IpFilterPolicy for secure public server

IpFilterPolicy
{
   PreDecap             off
   IpFilterGroupRef     ZoneB
}

KeyExchangePolicy
{
   AllowNat             Yes
   NatKeepAliveInterval 0
   KeyExchangeRuleRef   ZoneB_KeyExRule1
```

```
}

###### All re-usable statements follow #######
IpFilterGroup           ZoneB
{
   IpFilterRuleRef      ZoneB-Permitted-traffic
   IpFilterRuleRef      FTPServer-ZoneB  #IPSec-protected
   IpFilterRuleRef      EE-ZoneB         #IPSec-protected
}

#########################################
# IpFilterRules                         #
#   defines:                            #
#      data endpoints                   #
#      Allowed services                 #
#      Actions (permit, deny, ipsec)    #
#########################################
IpFilterRule            ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              IKE-local-500
   IpServiceRef              IKE-local-4500
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}

IpFilterRule           EE-ZoneB
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceRef              Enterprise-Extender
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         EE-vpnaction
   IpLocalStartActionRef     ZoneB-Start-Action
}

IpFilterRule           FTPServer-ZoneB
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrSetRef          ZoneB-subnet
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         FTP-vpnaction
}

#########################
# Local Start Actions #
#########################
IpLocalStartAction           ZoneB-Start-Action
{
   AllowOnDemand             yes
   LocalPortGranularity      packet
   RemotePortGranularity     packet
   ProtocolGranularity       packet
   LocalIpGranularity        packet
   RemoteIpGranulatiry       packet
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
}

#######################
# IpService groups #
#######################
IpServiceGroup            FTPServer
{
   IpServiceRef           FTPServer-Control
   IpServiceRef           FTPServer-Data-Passive
}

IpServiceGroup            SecureFTPServer
{
```

```
   IpServiceRef           SecureFTPServer-Control
   IpServiceRef           SecureFTPServer-Data-Passive
}

#########################################
# Services provided by this host #
#########################################

IpService                    IKE-local-500
{
  SourcePortRange            500
  DestinationPortRange       500
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    IKE-local-4500
{
  SourcePortRange            4500
  DestinationPortRange       4500
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}


IpService                    SecureFTPServer-Control
{
  SourcePortRange            990
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    SecureFTPServer-Data-Passive
{
  SourcePortRange            50201 50400
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    Enterprise-Extender
{
  SourcePortRange            12000 12004
  DestinationPortRange       12000 12004
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    FTPServer-Control
{
  SourcePortRange            21
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                  FTPServer-Data-Passive
  {
    SourcePortRange            50000 50200
```

```
    Protocol                    tcp
    Direction                   bidirectional InboundConnect
    Routing                     local
    SecurityClass               0
  }

##########################
# Security Endpoints #
##########################
LocalSecurityEndpoint  Local_IKED
{
   Identity      Fqdn  Server.PCO1.example.com
   Location      10.1.1.1
   CaLabel       CA4PartnerCompany
}RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      Fqdn   *.PCO2.example.com
   Location      9.4.0.0/16
}

#############################
# Generic filter actions #
#############################

IpGenericFilterAction    permit-nolog
{
   IpFilterAction       permit
   IpFilterLogging       no
}

IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction       ipsec
   IpFilterLogging       no
}

#####################################
# Key Exchange offers            #
#   defines:                     #
#     Authentication type        #
#     Encryption type            #
#     Peer authentication method #
#     Refresh limits             #
#####################################
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RsaSignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

#####################################
# Key Exchange Actions           #
#  defines:                      #
#    Negotiation mode            #
#    List of Key exchange offers #
#####################################
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate       main
   HowToRespondIKEv1   main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}

#############################################
# KeyExchangeRules                    #
#   defines:                          #
```

```
#      A pair of security endpoints  #
#      permitted in IKE negotiations #
########################################
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}

################################
# Data Offers                  #
#   defines:                   #
#      Encapsulation mode      #
#      Authentication type     #
#      Encryption type         #
#      Refresh limits          #
################################
### Authenticated offer ###
IpDataOffer TRAN-ESPSHA-NOENCR
{
   HowToEncap Transport
   HowToEncrypt DoNot
   HowToAuth    ESP HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

### Encrypted Authenticated offer ###
IpDataOffer TRAN-ESPSHA-3DES
{
   HowToEncap Transport
   HowToEncrypt 3DES
   HowToAuth    ESP HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

################################
# Dynamic VPN Actions          #
#   defines:                   #
#     Initiation role          #
#     Pfs group                #
#     Lifetime of connection   #
#     List of Data offers      #
################################
IpDynVpnAction FTP-vpnaction
{
   Initiation           remoteonly
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-ESPSHA-NOENCR
}

IpDynVpnAction EE-vpnaction
{
   Initiation           localonly
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-ESPSHA-3DES
}

##################
# IP addresses #
##################
```

```
IpAddr    PrivateServerAddress
{
   Addr   10.1.1.1
}IpAddrSet ZoneB-subnet
{
   Prefix 9.4.0.0/16
}
```

Steps for configuring the partner company with NAPT model (host-to-host with IPSec)

The partner company with NAPT model modifies the partner company with NAT model, replacing the NAT device in front of the internal network of each partner company with a NAPT device.

Before you begin

The following statements and concepts are covered in the discussion of this model:

- Using wildcard values for the remote port for IKE traffic
- NAPT implications for host-to-host dynamic IKE negotiations:
  - Local and remote data endpoints
  - Local and remote security endpoints
  - IKE initiator and responder roles

The partner company model assumed a network topology with both partner companies using public IP addresses in their internal networks. The partner company with NAT model modified the partner company model to include private addressing in the private network of each partner company, with a NAT device in front of each private network. Both NAT devices used static one-to-one address mappings.

The partner company with NAPT model modifies the partner company with NAT model, replacing the NAT device in front of the partner company's internal network with a NAPT device. The NAPT device uses many-to-one address and port mappings. The NAT in front of the z/OS host continues to use static one-to-one address mappings.

Figure 127 on page 1012 shows the partner company with NAPT topology.

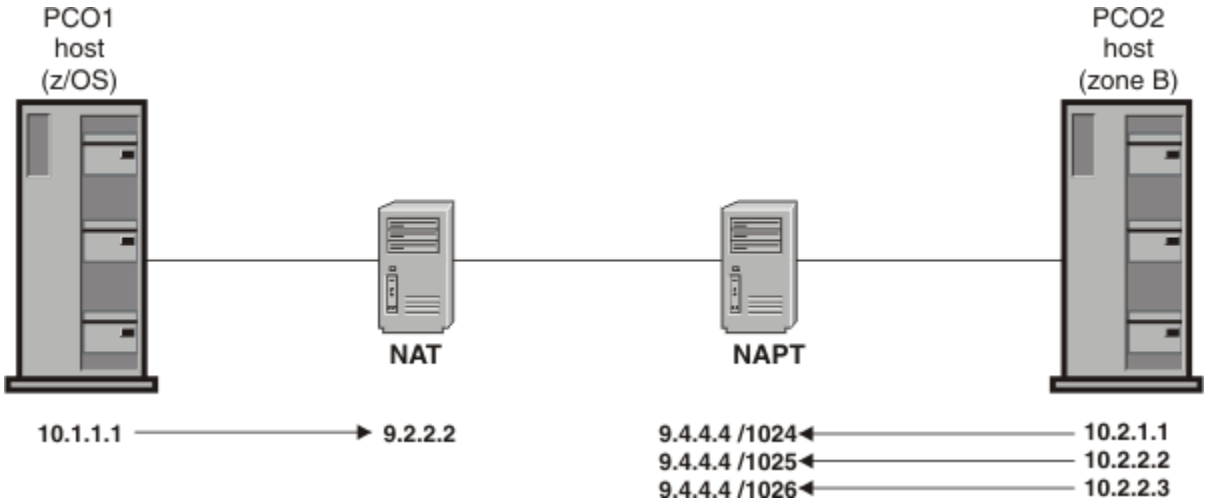


Figure 127. Partner company with NAPT model

The partner company with NAPT model has the same basic security requirements as the partner company with NAT model. One difference is that IPSec protection for Enterprise Extender (EE) traffic has been eliminated from the example. EE traffic is not compatible with the use of NAPT, with or without IPSec protection. For more information, see "Enterprise Extender considerations when traversing a NAT" on page 1040.

This example includes only the configuration steps that are impacted by a change from static one-to-one address mappings to NAPT mappings. Configuration statements added or changed for the partner company with NAPT model are shown in **bold**. The example describes the policy for partner company 1 (PCO1).

For this example, assume you must meet the following requirements to enable network communications from a partner company (PCO2) in an untrusted zone B behind a NAPT (9.4.4.4), over a connected network, to a server on this host that is behind a NAT, using static one-to-one address mapping:

- IKE traffic from untrusted zone B is allowed to this host.
- Secure FTP traffic (using TLS/SSL) from untrusted zone B is allowed to a secure FTP server running on this host.
- FTP traffic from untrusted zone B is allowed to an FTP server running on this host using a dynamic IPSec tunnel with strong authentication.
- A dynamic IPSec tunnel for normal FTP control activates for each remote host that initiates an FTP connection (remote activation).
- A dynamic IPSec tunnel for normal FTP data activates for each remote host that initiates an FTP data connection (remote activation).
- Peers authenticate themselves using the RSA signature method.

Procedure

Starting with the partner company with NAT policy, the following changes need to be made to meet these requirements when there is a NAPT in front of the partner company's internal network (Zone B):

- Modify the IpService statements for the IKE traffic to specify a wildcard value for the remote port.

  The NAPT will update both the IP address and port values of an IKE packet, so that the traffic can no longer be described with a constant value of 500 or 4500.

```
IpService                       IKE-local-500
{
  SourcePortRange               500
  DestinationPortRange          0
  Protocol                      UDP
  Direction                     bidirectional
  Routing                       local
  SecurityClass                 0
}

IpService                       IKE-local-4500
{
  SourcePortRange               4500
  DestinationPortRange          0
  Protocol                      UDP
  Direction                     bidirectional
  Routing                       local
  SecurityClass                 0
}
```

- Modify the remote data endpoint to reflect the NAPT's single IP address, 9.4.4.4.

  The remote data endpoint is the partner company's internal network. The z/OS implementation does not require you to code private addresses for the remote endpoint; the network address translated public address should be configured as the remote data endpoint. In this example, the private addresses in the PCO2 internal network (10.2.0.0/16) are translated into the public address 9.4.4.4. Thus, the remote data endpoint is 9.4.4.4.

```
IpAddr    ZoneB
{
   Addr   9.4.4.4
}
```

- Modify the remote security endpoint.

```
RemoteSecurityEndpoint   ZoneB_IKED
{
```

```
   Identity      Fqdn  *.PCO2.example.com
   Location      9.4.4.4
   CaLabel       CA4PartnerCompany
}
```

Results

A completely configured policy, including all objects and their references, is as follows:

```
# IpFilterPolicy for secure public server

IpFilterPolicy
{
   PreDecap             off
   IpFilterGroupRef     ZoneB
}

KeyExchangePolicy
{
   AllowNat             Yes
   NatKeepAliveInterval 0
   KeyExchangeRuleRef   ZoneB_KeyExRule1
}

###### All re-usable statements follow #######
IpFilterGroup           ZoneB
{
   IpFilterRuleRef      ZoneB-Permitted-traffic
   IpFilterRuleRef      FTPServer-ZoneB  #IPSec-protected
}

#########################################
# IpFilterRules                         #
#   defines:                            #
#      data endpoints                   #
#      Allowed services                 #
#      Actions (permit, deny, ipsec)    #
#########################################
IpFilterRule            ZoneB-Permitted-traffic
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrRef             ZoneB
   IpServiceRef              IKE-local-500
   IpServiceRef              IKE-local-4500
   IpServiceGroupRef         SecureFTPServer
   IpGenericFilterActionRef  permit-nolog
}

IpFilterRule           FTPServer-ZoneB
{
   IpSourceAddrRef           PrivateServerAddress
   IpDestAddrRef             ZoneB
   IpServiceGroupRef         FTPServer
   IpGenericFilterActionRef  ipsec-nolog
   IpDynVpnActionRef         FTP-vpnaction
}

######################
# IpService groups #
######################
IpServiceGroup            FTPServer
{
   IpServiceRef           FTPServer-Control
   IpServiceRef           FTPServer-Data-Passive
}

IpServiceGroup            SecureFTPServer
{
   IpServiceRef           SecureFTPServer-Control
   IpServiceRef           SecureFTPServer-Data-Passive
}

#####################################
# Services provided by this host #
#####################################

IpService                    IKE-local-500
{
  SourcePortRange            500
```

```
  DestinationPortRange       0
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    IKE-local-4500
{
  SourcePortRange            4500
  DestinationPortRange       0
  Protocol                   UDP
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}

IpService                    SecureFTPServer-Control
{
  SourcePortRange            990
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    SecureFTPServer-Data-Passive
{
  SourcePortRange            50201 50400
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                    FTPServer-Control
{
  SourcePortRange            21
  DestinationPortRange       1024 65535
  Protocol                   tcp
  Direction                  bidirectional InboundConnect
  Routing                    local
  SecurityClass              0
}

IpService                  FTPServer-Data-Passive
  {
    SourcePortRange            50000 50200
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }

##########################
# Security Endpoints #
##########################
LocalSecurityEndpoint  Local_IKED
{
   Identity      Fqdn  Server.PCO1.example.com
   Location      10.1.1.1
}

RemoteSecurityEndpoint   ZoneB_IKED
{
   Identity      Fqdn   *.PCO2.example.com
   Location      9.4.4.4
   CaLabel       CA4PartnerCompany
}

#############################
# Generic filter actions #
#############################

IpGenericFilterAction    permit-nolog
{
   IpFilterAction        permit
   IpFilterLogging       no
}
```

```
IpGenericFilterAction    ipsec-nolog
{
   IpFilterAction        ipsec
   IpFilterLogging       no
}

#######################################
# Key Exchange offers                 #
#   defines:                          #
#     Authentication type             #
#     Encryption type                 #
#     Peer authentication method      #
#     Refresh limits                  #
#######################################
KeyExchangeOffer RSA-SHA1-3DES-DH2
{
   HowToEncrypt    3DES
   HowToAuthMsgs   SHA1
   HowToAuthPeers  RsaSignature
   DHGroup         Group2
   RefreshLifetimeProposed  480
   RefreshLifetimeAccepted  240 1440
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

#######################################
# Key Exchange Actions                #
#  defines:                           #
#    Negotiation mode                 #
#    List of Key exchange offers      #
#######################################
KeyExchangeAction Main-RSA-SHA1-3DES-DH2
{
   HowToInitiate       main
   HowToRespondIKEv1   main
   KeyExchangeOfferRef RSA-SHA1-3DES-DH2
}

#############################################
# KeyExchangeRules                          #
#   defines:                                #
#      A pair of security endpoints         #
#      permitted in IKE negotiations        #
#############################################
KeyExchangeRule              ZoneB_KeyExRule1
{
   LocalSecurityEndpointRef  Local_IKED
   RemoteSecurityEndpointRef ZoneB_IKED
   KeyExchangeActionRef      Main-RSA-SHA1-3DES-DH2
}

###############################
# Data Offers                 #
#   defines:                  #
#      Encapsulation mode     #
#      Authentication type    #
#      Encryption type        #
#      Refresh limits         #
###############################
### Authenticated offer ###
IpDataOffer TRAN-ESPSHA-NOENCR
{
   HowToEncap   Transport
   HowToEncrypt DoNot
   HowToAuth    ESP HMAC_SHA1
   RefreshLifetimeProposed 240
   RefreshLifetimeAccepted 120 480
   RefreshLifesizeProposed  none
   RefreshLifesizeAccepted  none
}

###############################
# Dynamic VPN Actions         #
#   defines:                  #
#     Initiation role         #
#     Pfs group               #
#     Lifetime of connection  #
#     List of Data offers     #
###############################
IpDynVpnAction FTP-vpnaction
{
```

```
   Initiation           remoteonly
   InitiateWithPfs      group2
   AcceptablePfs        group2
   VpnLife              1440
   IpDataOfferRef       TRAN-ESPSHA-NOENCR
}


##################
# IP addresses #
##################

IpAddr    PrivateServerAddress
{
   Addr   10.1.1.1
}

IpAddr    ZoneB
{
   Addr   9.4.4.4
}
```

Steps for configuring the branch office model: Part 1 (host-to-gateway with IPSec)

The branch office model part 1 consists of two networks whose IP connectivity relies on the Internet. The server is protecting traffic that originates from hosts outside the internal network, which at some point is routed over the Internet.

Before you begin

The following topics are covered in the discussion of this model:

- Host-to-gateway
- Gateway-to-gateway
- Autoactivation
- Command-line activation
- Using wildcards for location and identity
- Pre-shared key peer authentication

Figure 128 on page 1017 shows the branch office portion of the security model network.

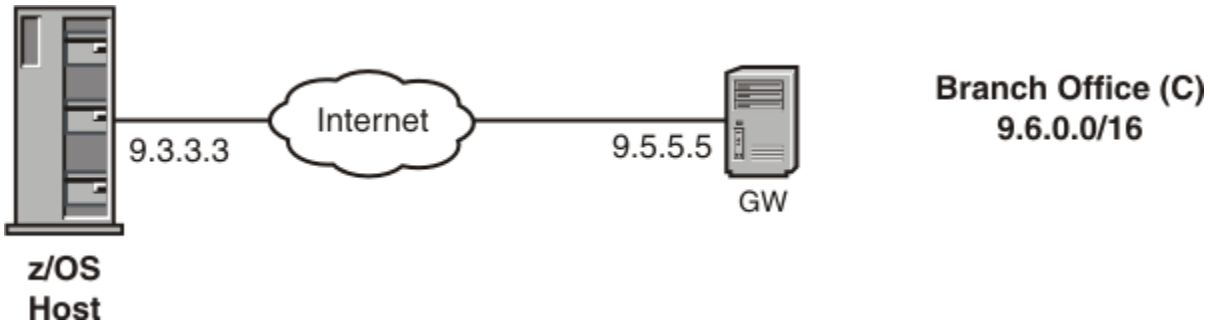


Figure 128. Branch office model

Transport-mode IPSec is used exclusively for transporting encrypted data directly between two hosts. Tunnel-mode IPSec encapsulation must be used if one of the security endpoints is a security gateway routing traffic for any number of hosts. In the branch office model, there can be multiple hosts behind a security gateway that is providing the security mechanism for all hosts that are behind the security gateway. The base assumption is that the local secure server represents a single data endpoint and a single security endpoint, whereas the remote subnetwork represents multiple data endpoints which share a common security endpoint, namely the security gateway host. Traffic from the local server to the branch office gateway is IPSec protected, while traffic from the branch office gateway to the hosts behind the security gateway need not be encrypted or even filtered.

As in the previous models, a complete IP security policy defines the traffic that is allowed between the local server and the zone representing the branch office. However, the difference lies primarily in where the remote security endpoint is situated. In the previous examples, all of the IPSec protection was provided on a host-to-host basis. Each set of communicating endpoints had a single dynamic VPN that represented a secure channel of communication between two hosts, local and remote. In contrast, any scenario that involves an IPSec security gateway can require that one VPN carry traffic for multiple hosts. This difference is highlighted in this branch office example.

For this example, assume the following requirements to allow network communications from zone C, a branch office network (9.6.0.0/16), to a public IP address (9.3.3.3) on this host. The hosts on the branch office network connect to the Internet through the public branch office gateway server (9.5.5.5).

- Allow IKE traffic from branch office zone C to this host.
- Allow EE traffic from branch office zone C to an EE service running on this host, using a dynamic VPN with strong authentication and encryption. The VPN should be up when the stack initializes.
- Allow normal FTP traffic from branch office zone C to an FTP server running on this host, using a dynamic VPN with strong authentication and encryption. The VPN is command-line activated. Only one VPN should be established to carry all FTP traffic. There will not be one VPN per connection, but rather one Security Association will be negotiated for all of the remote FTP clients.

  Restriction: Negotiating a single Security Association for multiple remote clients is possible only when the remote security endpoint is acting as a secure gateway.
- Peers authenticate themselves using the pre-shared key method.

Procedure

Perform the following steps to meet these requirements and configure part 1 of the branch office model.

1. Determine the number of zones to be protected.

   The branch office represents one zone, zone C.
2. For each zone, determine what services are allowed and define an IpService block for each wanted service.

   Services are defined by their protocols and the well-known ports they use.

   The definitions that describe FTP traffic can be combined in an IP service group as follows:

```
IpServiceGroup               FTPServer
{
  IpService                  FTPServer-Control
  {
    SourcePortRange            21
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }
  IpService                  FTPServer-Data
  {
    SourcePortRange            20
    Protocol                   tcp
    Direction                  bidirectional OutboundConnect
    Routing                    local
    SecurityClass              0
  }
  IpService                  FTPServer-Data-Passive
  {
    SourcePortRange            50000 50200
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }
}
```

The traffic pattern for Enterprise Extender can be defined in one IpService block as follows:

```
IpService                    Enterprise-Extender
{
  SourcePortRange            12000 12004
  DestinationPortRange       12000 12004
  Protocol                   udp
  Direction                  bidirectional
  Routing                    local
  SecurityClass              0
}
```

3. Determine the data endpoints to be protected.

   Typically, this is both a local and remote IP address or subnetwork.

   In this example, for the local host public IP address, 9.3.3.3, define the following statement:

```
IpAddr                       PublicServerAddressA1
{
  Addr                       9.3.3.3
}
```

   For the remote subnet, 9.6.0.0/16, the following statement is defined:

```
IpAddrSet                    SubnetC
{
  Prefix                     9.6.0.0/16
}
```

   Because it is necessary to permit IKE traffic between the local public server and the remote branch office gateway, the IP address of the remote gateway must also be defined as follows:

```
IpAddr                       BranchOfficeGateway
{
  Addr                       9.5.5.5
}
```

4. Determine what level of security is needed between each set of data endpoints.

   In this example, strong authentication and encryption is needed for both EE and FTP traffic, so ESP authentication and ESP encryption are used.

5. Configure an IpGenericFilterAction statement for the level of security that is required (permit, deny, ipsec), including whether the connection should be logged, as follows:

```
IpGenericFilterAction        ipsec
{
  IpFilterAction             ipsec
}
```

6. If IPSec is required between any two endpoints, take the following actions:

   a) Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:

      i) Determine the required type and strength of protection for the phase 1 Security Association.

      In this example, SHA1 authentication is used because it is more secure than MD5, and AES encryption is used because it is stronger than DES. Diffie-Hellman Group5 and Group14 are considered strong enough to generate keying material for AES using a 128 bit key. Group14 is used in this example.

      ii) Decide what type of peer authentication to use.

      In this example, pre-shared key authentication is specified in the requirements. Typically, the RSA signature method is preferable, given its numerous advantages, but for the purposes of example the pre-shared key method is used here. Because there is only one remote IKE peer in the branch office scenario (the remote security gateway), and because it is relatively simple to configure, pre-shared key authentication is a reasonable choice.

iii) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation as follows:

```
KeyExchangeOffer             SHA1-AES-PSK
{
  HowToEncrypt               AES_CBC KeyLength 128
  HowToAuthMsgs              SHA1
  HowToAuthPeers             PresharedKey
  DHGroup                    Group14
}
```

iv) Decide whether NAT traversal will be allowed.

In this example, the following parameter is used:

```
AllowNat No
```

v) Determine the negotiation mode, Main or Aggressive.

Because security is a priority in the branch office model, the more secure Main mode is used for the phase 1 negotiation.

vi) Configure a KeyExchangeAction statement as follows:

```
KeyExchangeAction            Gold-PSK
{
  HowToInitiate              main
  HowToRespondIKEv1          main
  KeyExchangeOfferRef        SHA1-AES-PSK
}
```

vii) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement as follows:

```
LocalSecurityEndpoint        Public_IKED
{
  Identity                   IpAddr 9.3.3.3
  LocationRef                PublicServerAddressA1
}

RemoteSecurityEndpoint       ZoneC_IKED
{
  Identity                   Fqdn gateway.BO.example.com
  LocationRef                BranchOfficeGateway
}
```

viii) Configure a KeyExchangeRule statement that includes the two endpoints, the key exchange action, and the pre-shared key as follows:

```
KeyExchangeRule              ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef   Public_IKED
  RemoteSecurityEndpointRef  ZoneC_IKED
  KeyExchangeActionRef       Gold-PSK
  PresharedKey               abracadabra
}
```

ix) Include the KeyExchangeRule statement in the KeyExchangePolicy statement block, as follows:

```
KeyExchangePolicy
{
    KeyExchangeRuleRef         ZoneC_KeyExRule1
}
```

b) Configure an IpDynVpnAction statement defining the control of the phase 2 negotiation, as follows:

i) Determine the required type and strength of IPSec protection for the phase 2 Security Association.

In this example, authentication is ESP HMAC_SHA1 and encryption is AES.

ii) Determine whether tunnel or transport mode is required.

Tunnel mode is required when one of the security endpoints is a security gateway.

iii) Configure an IpDataOffer statement that defines the parameters of the phase 2 negotiation, as follows:

```
IpDataOffer                  SHA-AES-Tunnel
{
  HowToEncap                 tunnel
  HowToEncrypt               AES_CBC KeyLength 128
  HowToAuth                  ESP HMAC_SHA1
}
```

iv) Determine which peer is allowed to initiate the negotiation.

Because the EE VPN is up when the stack starts, you need to be able to start the negotiation locally (that is, autoactivate it). The FTP VPN is command-line activated, so you need to be able to start the negotiation locally.

v) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation, as follows:

```
IpDynVpnAction               Gold-TunnelMode
{
  Initiation                 either
  InitiateWithPfs            group2
  AcceptablePfs              group2
  IpDataOfferRef             SHA-AES-Tunnel
}
```

c) Decide how the Security Association is activated.

i) Optionally, configure a local dynamic VPN policy, if command-line activated or autoactivated.

A LocalDynVpnRule statement is required for each Security Association because none are on-demand activated, five for the EE traffic and two for the FTP traffic, as follows:

```
LocalDynVpnRule              ZoneC_VPN-EE1
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              12000
   RemoteDataPort             12000
   Protocol                   UDP
   AutoActivate               Yes
}

LocalDynVpnRule              ZoneC_VPN-EE2
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              12001
   RemoteDataPort             12001
   Protocol                   UDP
   AutoActivate               Yes
}

LocalDynVpnRule              ZoneC_VPN-EE3
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              12002
   RemoteDataPort             12002
   Protocol                   UDP
   AutoActivate               Yes
```

```
}

LocalDynVpnRule              ZoneC_VPN-EE4
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              12003
   RemoteDataPort             12003
   Protocol                   UDP
   AutoActivate               Yes
}

LocalDynVpnRule              ZoneC_VPN-EE5
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              12004
   RemoteDataPort             12004
   Protocol                   UDP
   AutoActivate               Yes
}

LocalDynVpnRule              ZoneC_VPN-FTP-Data
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              20
   RemoteDataPort             0
   Protocol                   TCP
   AutoActivate               Yes
}

LocalDynVpnRule              ZoneC_VPN-FTP-Control
{
   LocalIpRef                 PublicServerAddressA1
   RemoteIpSetRef             SubnetC
   LocalDataPort              21
   RemoteDataPort             0
   Protocol                   TCP
   AutoActivate               Yes
}
```

ii) Configure an optional IpLocalStartAction statement, if the Security Association is to be activated locally (that is, on-demand, command-line, or autoactivation) and one of the security endpoints is acting as a security gateway (that is, not a host-to-host Security Association). Include a reference to the IpLocalStartAction statement in the IP filter rule.

If the local secure server initiates an IKE negotiation, it must be able to identify a remote IKE peer. However, in this case, the remote data endpoint is not a single host, but multiple endpoints in a subnetwork, ZoneC. The IpLocalStartAction statement is used to identify the remote IKE peer, and is required if the local IKE daemon initiates an ipsec connection with a remote security gateway.

```
IpLocalStartAction             StartZoneC
{
   RemoteSecurityEndpointRef   ZoneC_IKED
}
```

iii) Create an IpFilterRule statement that allows IPSec traffic (AH and ESP), or set the global IpFilterPolicy statement parameter PreDecap to `off`.

In this example, `PreDecap off` is used in the IpFilterPolicy statement.

7. Define an IpFilterRule statement for each set of data endpoints.

The rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port

500). If NAT traversal is allowed, create an IpFilterRule statement that allows IKE UDP traffic on port 4500. See the IpAddr and IpAddrSet statements configured in step "3" on page 1019.

```
IpFilterRule                 Rule1C
{
   IpSourceAddrRef            PublicServerAddressA1
   IpDestAddrRef              BranchOfficeGateway
   IpServiceRef               IKE
   IpGenericFilterActionRef   permit
}

IpFilterRule                 Rule2C
{
   IpSourceAddrRef            PublicServerAddressA1
   IpDestAddrSetRef           SubnetC
   IpServiceRef               Enterprise-Extender
   IpServiceGroupRef          FTPServer
   IpGenericFilterActionRef   ipsec
   IpDynVpnActionRef          Gold-TunnelMode
   IpLocalStartActionRef      StartZoneC
}
```

8. Define an IpFilterGroup statement for each zone and include the IpFilterRule statements that belong to that zone.
9. Include all IpFilterGroup references and, optionally, any additional IpFilterRule statements as needed, in the IpFilterPolicy block.
10. Include all configured statements in the stack-specific IP security configuration file.

Results

The following policy is a complete IP security policy for traffic from the local secure server to zone C, assuming that the reusable statements are included in the common IP security configuration file:

```
#---------------------------------------------------------
# Filter Policy for Secure Server
#---------------------------------------------------------
IpFilterPolicy
{
  PreDecap                   off
  FilterLogging              on
  AllowOnDemand              no
  IpFilterGroupRef           ZoneC
}

#---------------------------------------------------------
# KeyExchange Policy for Secure Server
#---------------------------------------------------------
KeyExchangePolicy
{
  KeyExchangeRuleRef         ZoneC_KeyExRule1
}

#---------------------------------------------------------
# LocalDynVpn Policy for Secure Server
#---------------------------------------------------------
LocalDynVpnPolicy
{
  LocalDynVpnGroupRef        ZoneC_BranchOfficeVPNs
}

##########################################################################
#               Connectivity Profile
#              Secure Server To Zone C
#
#   Server to Trusted Branch Office Network
#
##########################################################################
IpFilterGroup                ZoneC
```

```
{
   #-------------------------------------------------------
   # Permitted Zone C traffic:
   #    Allow IKE traffic from the gateway IKE Server
   #    for branch office to this host.
   #
   #    IKE  (UDP port 500) - IKE negotiations
   #-------------------------------------------------------

   IpFilterRule                Rule1C
   {
     IpSourceAddrRef           PublicServerAddressA1
     IpDestAddrRef             BranchOfficeGateway
     IpServiceRef              IKE
     IpGenericFilterActionRef  permit
   }

   #-------------------------------------------------------
   # IPSec-protected Zone C traffic:
   #
   # Enterprise Extender (ports 12000-12004)
   #    FTP Server - SubnetC to PublicServerAddressA
   #-------------------------------------------------------

   IpFilterRule                Rule2C
   {
     IpSourceAddrRef           PublicServerAddressA1
     IpDestAddrSetRef          SubnetC
     IpServiceRef              Enterprise-Extender
     IpServiceGroupRef         FTPServer
     IpGenericFilterActionRef  ipsec
     IpDynVpnActionRef         Gold-TunnelMode
     IpLocalStartActionRef     StartZoneC
   }
}

IpLocalStartAction            StartZoneC
{
   AllowOnDemand              yes
   RemoteSecurityEndpointRef  ZoneC_IKED
}

KeyExchangeRule             ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef  Public_IKED
  RemoteSecurityEndpointRef ZoneC_IKED
  KeyExchangeActionRef      Gold-PSK
}

#-------------------------------------------------------
# Zone C LocalDynVpnRules
#
# Setup SAs for EE traffic from branch office zone C to
# EE (UDP ports 12000-12004).
#-------------------------------------------------------

LocalDynVpnGroup            ZoneC_BranchOfficeVPNs
{
   LocalDynVpnRule             ZoneC_VPN-EE1
   {
     LocalIpRef                PublicServerAddressA1
     RemoteIpSetRef            SubnetC
     LocalDataPort             12000
     RemoteDataPort            12000
     Protocol                  UDP
     AutoActivate              Yes
   }

   LocalDynVpnRule             ZoneC_VPN-EE2
   {
     LocalIpRef                PublicServerAddressA1
     RemoteIpSetRef            SubnetC
```

```
      LocalDataPort              12001
      RemoteDataPort             12001
      Protocol                   UDP
      AutoActivate               Yes
   }

   LocalDynVpnRule               ZoneC_VPN-EE3
   {
      LocalIpRef                 PublicServerAddressA1
      RemoteIpSetRef             SubnetC
      LocalDataPort              12002
      RemoteDataPort             12002
      Protocol                   UDP
      AutoActivate               Yes
   }

   LocalDynVpnRule               ZoneC_VPN-EE4
   {
      LocalIpRef                 PublicServerAddressA1
      RemoteIpSetRef             SubnetC
      LocalDataPort              12003
      RemoteDataPort             12003
      Protocol                   UDP
      AutoActivate               Yes
   }

   LocalDynVpnRule               ZoneC_VPN-EE5
   {
      LocalIpRef                 PublicServerAddressA1
      RemoteIpSetRef             SubnetC
      LocalDataPort              12004
      RemoteDataPort             12004
      Protocol                   UDP
      AutoActivate               Yes
   }

#-------------------------------------------------------
# Setup SAs for FTP traffic from branch office zone C
# to an FTP server running on this host using a dynamic
# vpn (TCP port 20, 21).
#-------------------------------------------------------

   LocalDynVpnRule               ZoneC_VPN-FTP-Data
   {
      LocalIpRef                 PublicServerAddressA1
      RemoteIpSetRef             SubnetC
      LocalDataPort              20
      RemoteDataPort             0
      Protocol                   TCP
      AutoActivate               Yes
   }

   LocalDynVpnRule               ZoneC_VPN-FTP-Control
   {
      LocalIpRef                 PublicServerAddressA1
      RemoteIpSetRef             SubnetC
      LocalDataPort              21
      RemoteDataPort             0
      Protocol                   TCP
      AutoActivate               Yes
   }
}
```

Steps for configuring the branch office with NAT model (host-to-gateway with IPSec)

The branch office with NAT model modifies the branch office model topology to include a NAT in front of the security gateway.

Before you begin

The following NAT implications for host-to-gateway dynamic IKE negotiations are covered in the discussion of this model:

- Local and remote data endpoints
- Local and remote security endpoints
- IKE initiator and responder roles
- Restriction on HowToAuth protocol (AH not supported)

"Steps for configuring the branch office model: Part 1 (host-to-gateway with IPSec)" on page 1017 assumed a network topology with both the host and the security gateway, as well as the hosts behind the security gateway, using public IP addresses. Often one or both security endpoints are behind a NAT using a private IP address.

Modifying the branch office model topology to include a NAT in front of the security gateway, the branch office with NAT topology becomes as shown in Figure 129 on page 1026:

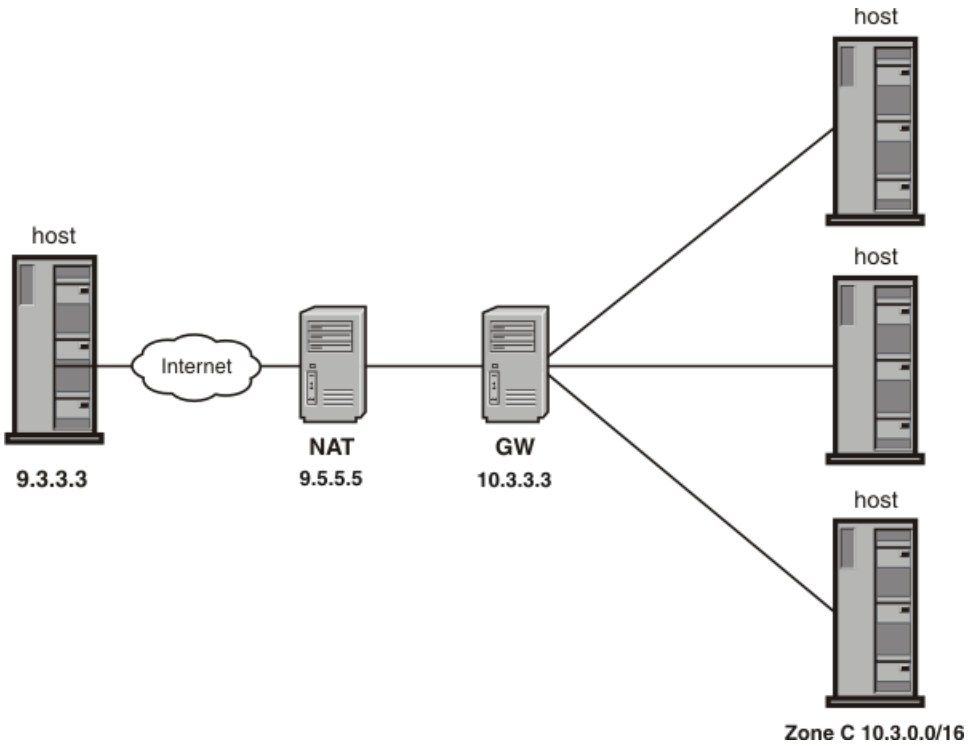


Figure 129. Branch office with NAT model

This example will describe the configuration considerations and requirements when the NATT solution is implemented to traverse NAT devices in a host-to-security gateway environment. The branch office with NAT model has the same basic security requirements as the branch office model. Configuration statements added or changed for the branch office with NAT model are shown in **bold**. The example describes the policy for host 9.3.3.3.

For this example, assume the following requirements to enable network communications from zone C, a branch office network using private IP addresses (10.3.0.0/16), to a public IP address (9.3.3.3) on this host. The hosts on the branch office network connect to the Internet through the branch office gateway

server, which is behind a NAT. In this model, the NAT has a static mapping of the security gateway's private address 10.3.3.3 to the public address 9.5.5.5.

- Allow IKE traffic from the branch office zone C security gateway to this host.
- Allow EE traffic from branch office zone C to an EE service running on this host, using a dynamic VPN with strong authentication and encryption. Only one host behind the security gateway (branch office zone C) will be able to send EE traffic.

  Guideline: In most cases, EE hosts should not be located behind a security gateway that is behind a NAT. Instead, a host-to-host Security Association should be negotiated for each EE host.

  In "Steps for configuring the branch office model: Part 1 (host-to-gateway with IPSec)" on page 1017, the VPN was brought up when the z/OS stack initialized by setting the AutoActivate parameter to Yes on the LocalDynVpnRule statement. In this scenario, the hosts behind the gateway do not have public addresses that can be configured in the policy. Therefore, initiation from host 9.3.3.3 to the security gateway 9.5.5.5 becomes ambiguous because the IP address of the remote data endpoint is unknown. z/OS does not allow initiation of a UDP-encapsulated tunnel mode Security Association to a security gateway. The Security Associations between the security gateway 9.5.5.5 and host 9.3.3.3 must be initiated by the security gateway, with host 9.3.3.3 acting as responder.
- Allow normal FTP traffic from branch office zone C to an FTP server running on this host, using a dynamic VPN with strong authentication and encryption. In "Steps for configuring the branch office model: Part 1 (host-to-gateway with IPSec)" on page 1017, the VPN was activated by an administrator from the z/OS UNIX command line. Again, z/OS does not allow initiation of a UDP-encapsulated tunnel mode Security Association to a security gateway. The Security Associations between the security gateway and host 9.3.3.3 must be initiated by the security gateway, with host 9.3.3.3 acting as the responder.
- Peers authenticate themselves using the pre-shared key method.

Procedure

Perform the following steps to meet these requirements and configure the branch office with NAT model.

1. Determine the number of zones to be protected.

   The branch office represents one zone, zone C.
2. For each zone, determine what services are allowed and define an IpService block for each wanted service.

   Services are defined by their protocols and the well-known ports that they use.

   The definitions that describe FTP traffic can be combined in an IP service group. Typically, FTP clients use passive mode (PASV) or extended passive mode (EPSV) to connect to the FTP server when the client is behind a NAT. For more information on active and passive mode FTP, see "Considerations for IPSec-encapsulated FTP traffic when traversing a NAT" on page 1039.

   The range of server ports specified for the data connection reflects the port range specified for PASSIVEDATAPORTS in the server's FTP.DATA file. For more information on PASSIVEDATAPORTS, see z/OS Communications Server: IP Configuration Reference.

   The following definitions show the services required for the server for PASV or EPSV:

```
IpServiceGroup               FTPServer
{
  IpService                  FTPServer-Control
  {
    SourcePortRange            21
    Protocol                   tcp
    Direction                  bidirectional InboundConnect
    Routing                    local
    SecurityClass              0
  }
  IpService                  FTPServer-Data-Passive
  {
    SourcePortRange            50000 50200
```

```
    Protocol                    tcp
    Direction                   bidirectional InboundConnect
    Routing                     local
    SecurityClass               0
  }
}
```

The traffic pattern for Enterprise Extender can be defined in one IpService block as follows:

```
IpService                     Enterprise-Extender
{
  SourcePortRange             12000 12004
  DestinationPortRange        12000 12004
  Protocol                    udp
  Direction                   bidirectional
  Routing                     local
  SecurityClass               0
}
```

3. Determine the data endpoints to be protected.

   In this example, for the local host public IP address, 9.3.3.3, define the following statement:

```
IpAddr                        PublicServerAddressA1
{
  Addr                        9.3.3.3
}
```

   In this case, the remote data endpoints are in the branch office's internal network. The z/OS NATT implementation does not require the coding of private addresses for the remote endpoints. Instead, the security gateway's public address is treated as the remote data endpoint. The NAT is using a static mapping for the security gateway, so the public address of the gateway is specified. If the NAT was using dynamic mappings, the range of public IP addresses to which the security gateway could be mapped would need to be included in this definition.

```
IpAddr                        BranchOfficeGateway
{
  Addr                        9.5.5.5
}
```

   This IP address is also needed to permit IKE traffic between the local public server and the remote branch office gateway.

4. Determine what level of security is needed between each set of data endpoints.

   In this example, strong authentication and encryption is needed for both EE and FTP traffic, so ESP authentication and ESP encryption are used. ESP must be used when NAT traversal support is being used.

5. Configure an IpGenericFilterAction statement for the level of security that is required (permit, deny, ipsec), including whether the connection should be logged, as follows:

```
IpGenericFilterAction         ipsec
{
  IpFilterAction              ipsec
}
```

6. If IPSec is required between any two endpoints, take the following actions:

   a) Configure a key exchange policy that defines the parameters of the phase 1 negotiation as follows:

      i) Determine the required type and strength of protection for the phase 1 Security Association.

         In this example, SHA1 authentication is used because it is more secure than MD5, and AES encryption is used.

      ii) Decide what type of peer authentication to use.

In this example, pre-shared key authentication is specified in the requirements. Typically, the RSA signature method is preferable, given its numerous advantages, but for the purposes of example the pre-shared key method is used here. Because there is only one remote IKE peer in the branch office scenario (the remote security gateway), and because it is relatively simple to configure, pre-shared key authentication is a reasonable choice.

iii) Configure a KeyExchangeOffer statement that defines the parameters for the phase 1 negotiation as follows:

```
KeyExchangeOffer              SHA1-AES-PSK
{
  HowToEncrypt                AES_CBC KeyLength 128
  HowToAuthMsgs               SHA1
  HowToAuthPeers              PresharedKey
  DHGroup                     Group14
}
```

iv) Decide whether NAT traversal will be allowed.

The AllowNat parameter on the KeyExchangePolicy and KeyExchangeAction statements enables the IKED to advertise NAT traversal support. For this model, allow NAT traversal support to be advertised for all phase 1 Security Associations by specifying Yes on the AllowNat parameter on the KeyExchangePolicy statement:

```
AllowNat Yes
```

For this model, use the NatKeepAliveInterval parameter default value, 20 seconds. When z/OS is behind a NAT, a NAT keep-alive timer is started, with the interval specified in NatKeepAliveInterval parameter on the KeyExchangePolicy statement. Because z/OS is not behind a NAT in this model, a NAT keep-alive is not kept regardless of the value specified or the default for NatKeepAliveInterval.

v) Determine the negotiation mode, Main or Aggressive.

Because security is a priority in the branch office with NAT model, the more secure Main mode is used for the phase 1 negotiation.

vi) Configure a KeyExchangeAction statement as follows:

```
KeyExchangeAction             Gold-PSK
{
  HowToInitiate               main
  HowToRespondIKEv1           main
  KeyExchangeOfferRef         SHA1-AES-PSK
}
```

vii) Configure a LocalSecurityEndpoint statement and RemoteSecurityEndpoint statement as follows:

```
LocalSecurityEndpoint         Public_IKED
{
  Identity                    IpAddr 9.3.3.3
  LocationRef                 PublicServerAddressA1
}

RemoteSecurityEndpoint        ZoneC_IKED
{
  Identity                    Fqdn gateway.BO.example.com
  LocationRef                 BranchOfficeGateway
}
```

viii) Configure a KeyExchangeRule statement that includes the two endpoints and the key exchange action, as follows:

```
KeyExchangeRule               ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef    Public_IKED
```

```
  RemoteSecurityEndpointRef  ZoneC_IKED
  KeyExchangeActionRef       Gold-PSK
  PresharedKey               abracadabra
}
```

ix) Include the KeyExchangeRule statement in the KeyExchangePolicy statement block, as follows:

```
KeyExchangePolicy
{
    AllowNat                   Yes
    KeyExchangeRuleRef         ZoneC_KeyExRule1
}
```

b) Configure an IpDynVpnAction statement defining the control of the phase 2 negotiation, as follows:

i) Determine the required type and strength of IPSec protection for the phase 2 Security Association.

In this example, authentication is ESP HMAC_SHA1 and encryption is AES.

ii) Determine whether tunnel or transport mode is required.

Tunnel mode is required when one of the security endpoints is a security gateway.

iii) Configure an IpDataOffer statement that defines the parameters of the phase 2 negotiation, as follows:

```
IpDataOffer                SHA-AES-Tunnel
{
  HowToEncap               tunnel
  HowToEncrypt             AES_CBC KeyLength 128
  HowToAuth                ESP HMAC_SHA1
}
```

iv) Determine which peer is allowed to initiate the negotiation.

In this scenario, the hosts behind the gateway do not have public addresses that can be configured in the policy. Therefore, initiation from host 9.3.3.3 to the security gateway 9.5.5.5 becomes ambiguous because the IP address of the remote data endpoint is unknown. z/OS does not allow initiation of a UDP-encapsulated tunnel mode Security Association to a security gateway. The Security Associations between the security gateway 9.5.5.5 and host 9.3.3.3 must be initiated by the security gateway, with host 9.3.3.3 acting as responder.

v) Configure an IpDynVpnAction statement that defines the control information for the phase 2 negotiation, as follows:

```
IpDynVpnAction             Gold-TunnelMode
{
  Initiation               remoteonly
  InitiateWithPfs          group2
  AcceptablePfs            group2
  IpDataOfferRef           SHA-AES-Tunnel
}
```

c) Decide how the Security Association is activated.

i) Only remote initiation of the Security Association is allowed, as specified for the Initiation parameter on the Gold-TunnelMode IpDynVpnAction statement. No LocalDynVpnRule or IpLocalStartAction statements are needed.

ii) Create an IpFilterRule statement that allows IPSec traffic (ESP), or set the global IpFilterPolicy statement parameter PreDecap to `off`.

In this example, `PreDecap off` is used in the IpFilterPolicy statement.

7. Define an IpFilterRule statement for each set of data endpoints.

The rule should include the services that are allowed (one IpService statement for each allowed service), and the level of security that is required (a reference to the IpGenericFilterAction statement). If IPSec is required, create an IpFilterRule statement that allows IKE traffic (UDP, port 500 and 4500). See the IpAddr and IpAddrSet statements configured previously.

```
IpFilterRule                 Rule1C
{
   IpSourceAddrRef            PublicServerAddressA1
   IpDestAddrRef              BranchOfficeGateway
   IpServiceRef               IKE-local-500
   IpServiceRef               IKE-local-4500
   IpGenericFilterActionRef   permit
}

IpFilterRule                 Rule2C
{
   IpSourceAddrRef            PublicServerAddressA1
   IpDestAddrRef              BranchOfficeGateway
   IpServiceRef               Enterprise-Extender
   IpServiceGroupRef          FTPServer
   IpGenericFilterActionRef   ipsec
   IpDynVpnActionRef          Gold-TunnelMode
}
```

8. Define an IpFilterGroup statement for each zone and include the IpFilterRule statements that belong to that zone.
9. Include all IpFilterGroup references and, optionally, any additional IpFilterRule statements as needed, in the IpFilterPolicy block.
10. Include all configured statements in the stack-specific IP security configuration file.

Results

The following policy is the complete IP security policy for traffic from the local secure server to zone C, assuming that the reusable statements are included in the common IP security configuration file:

```
#---------------------------------------------------------
# Filter Policy for Secure Server
#---------------------------------------------------------
IpFilterPolicy
{
  PreDecap                   off
  FilterLogging              on
  AllowOnDemand              no
  IpFilterGroupRef           ZoneC
}

#---------------------------------------------------------
# KeyExchange Policy for Secure Server
#---------------------------------------------------------
KeyExchangePolicy
{
  AllowNat                   Yes
  KeyExchangeRuleRef         ZoneC_KeyExRule1
}

###########################################################################
#               Connectivity Profile
#              Secure Server To Zone C
#
#   Server to Trusted Branch Office Network
#
###########################################################################
IpFilterGroup                ZoneC
{
   #-------------------------------------------------------
   # Permitted Zone C traffic:
   #    Allow IKE traffic from the gateway IKE Server
   #    for branch office to this host.
```

```
   #
   #    IKE  (UDP port 500/4500) - IKE negotiations
   #------------------------------------------------------

   IpFilterRule                 Rule1C
   {
     IpSourceAddrRef            PublicServerAddressA1
     IpDestAddrRef              BranchOfficeGateway
     IpServiceGroupRef          IKE
     IpGenericFilterActionRef   permit
   }

   #------------------------------------------------------
   # IPSec-protected Zone C traffic:
   #
   #    Enterprise Extender (ports 12000-12004)
   #    FTP Server - SubnetC to PublicServerAddressA
   #------------------------------------------------------

   IpFilterRule                 Rule2C
   {
     IpSourceAddrRef            PublicServerAddressA1
     IpDestAddrRef              BranchOfficeGateway
     IpServiceRef               Enterprise-Extender
     IpServiceGroupRef          FTPServer
     IpGenericFilterActionRef   ipsec
     IpDynVpnActionRef          Gold-TunnelMode
   }
}


KeyExchangeRule              ZoneC_KeyExRule1
{
  LocalSecurityEndpointRef   Public_IKED
  RemoteSecurityEndpointRef  ZoneC_IKED
  KeyExchangeActionRef       Gold-PSK
  PresharedKey               abracadabra
}
```

Steps for configuring the branch office model: Part 2 (gateway-to-gateway with IPSec)

The branch office model part 2 is similar to the branch office model part 1. Here, there are multiple data endpoints on both the local side and the remote side, but only one pair of security endpoints, one local and one remote.

Before you begin

It is not likely that the z/OS system will function strictly as a router or a firewall at the network perimeter, but it is possible to configure the z/OS system to provide the IPSec functionality that many secure gateway devices provide. This topic includes instructions on how to configure a scenario in which the z/OS system is routing network traffic from inside the internal network. This functionality is similar to the functionality that is provided by the branch office gateway in “Steps for configuring the branch office model: Part 1 (host-to-gateway with IPSec)” on page 1017. Here, there are multiple data endpoints on both the local side and the remote side, but only one pair of security endpoints, one local and one remote.

In this example, assume that the local z/OS system is acting as a secure gateway for hosts on an internal network A, and tunneling the IPSec-protected traffic to a remote secure gateway for subnetwork C. The following list summarizes the requirements for this example:

- Permit IKE negotiations between the two security gateways, the secure local host and the secure remote gateway for subnetwork C.
- Permit traffic from the internal network to the internal interface on the secure local host.
- Add IPSec protection to any traffic that flows between the two secure gateways.

In this scenario, the z/OS system is a secure forwarding agent for the internal hosts, rather than a data endpoint. Traffic from the internal hosts that is destined for the remote network first comes to the secure local gateway in the clear. Before it is sent out to the remote network, it is IPSec encapsulated. The process is reversed for traffic that comes from the remote network. Traffic that comes from the remote network to the local secure gateway is IPSec decapsulated on the local secure host and forwarded to the internal host in the clear.

Procedure

Perform the following steps to meet the above requirements and configure part 2 of the branch office model.

1. Permit IKE negotiations between the two secure gateways.

   UDP port 500 traffic must be allowed for IKE negotiations.

```
IpFilterRule                  Rule1AtoC
{
     IpSourceAddrRef            PublicServerAddressA1
     IpDestAddrRef              BranchOfficeGateway
     IpServiceRef               IKE
     IpGenericFilterActionRef   permit
}
```

2. Permit traffic from the internal network to the internal interface on the secure host.

```
IpFilterRule                  Rule2AtoC
{
     IpSourceAddrSetRef         SubnetC
     IpDestAddrSetRef           InternalNetworkA
     IpServiceGroupRef          All-traffic-routed
     IpGenericFilterActionRef   permit
}
```

   The bidirectional keyword on the IpService statement creates two filter rules, one inbound and one outbound. Expansion of this IpFilterRule statement is shown in Table 51 on page 1033.

Table 51. Expanded filter rule for internal traffic

Source	Destination	Routing	Direction	Action
SubnetC	InternalNetworkA	Routed	Outbound	permit
InternalNetworkA	SubnetC	Routed	Inbound	permit

   As required, traffic that enters the secure server from InternalNetworkA that is destined for SubnetC is permitted by the secure host as an inbound routed packet. Traffic that leaves the secure server from SubnetC destined for InternalNetworkA is permitted by the secure host as an outbound routed packet.

3. Add IPSec protection to any traffic that flows between the two secure gateways.

```
IpFilterRule                 Rule3AtoC
{
     IpSourceAddrSetRef         InternalNetworkA
     IpDestAddrSetRef           SubnetC
     IpServiceGroupRef          All-traffic-routed
     IpGenericFilterActionRef   ipsec-log
}
```

   Expansion of this rule is shown in Table 52 on page 1033.

Table 52. Expanded filter rule for remote traffic

Source	Destination	Routing	Direction	Action
InternalNetworkA	SubnetC	Routed	Outbound	ipsec
SubnetC	InternalNetworkA	Routed	Inbound	ipsec

Traffic that leaves the secure server from InternalNetworkA that is destined for SubnetC is permitted with ipsec. Traffic that enters the secure server from SubnetC that is destined for InternalNetworkA is permitted with ipsec.

Additional topologies

There are alternative security models that might be suitable for particular networks. Each of these configurations is supported by z/OS IP security.

Cascaded tunnels

If there are multiple hops from data endpoint to data endpoint, there might be Security Associations between any two hosts along the path. For example, the data might be authenticated from a host to the secure gateway, then encrypted for transportation over the Internet, then possibly authenticated and encrypted from the second secure gateway to the host on the other side, as shown in Figure 130 on page 1034:

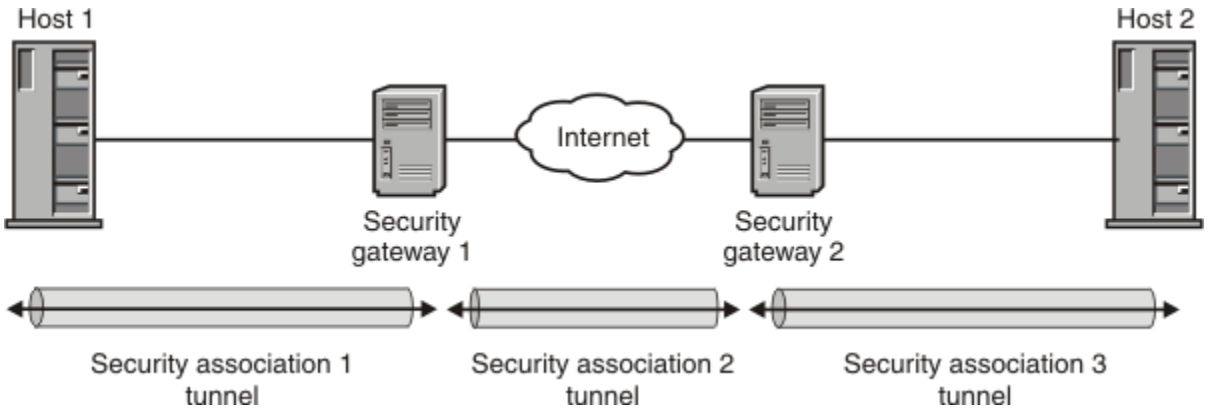


Figure 130. Cascaded tunnels

Nested tunnels

In a nested environment, data is encapsulated multiple times over multiple hops. If the local and remote hosts are both behind a secure gateway, there could be a tunnel-mode Security Association that carries the traffic from one secure gateway to the other. Meanwhile, a transport-mode Security Association could carry the traffic from one host to the other, end-to-end. In this case, the transport-mode Security Association is nested in the tunnel-mode Security Association. Data from the local host is encapsulated once when leaving the local machine, encapsulated again at the secure gateway, decapsulated at the other secure gateway, and decapsulated one final time at the remote host, leaving only the original packet.

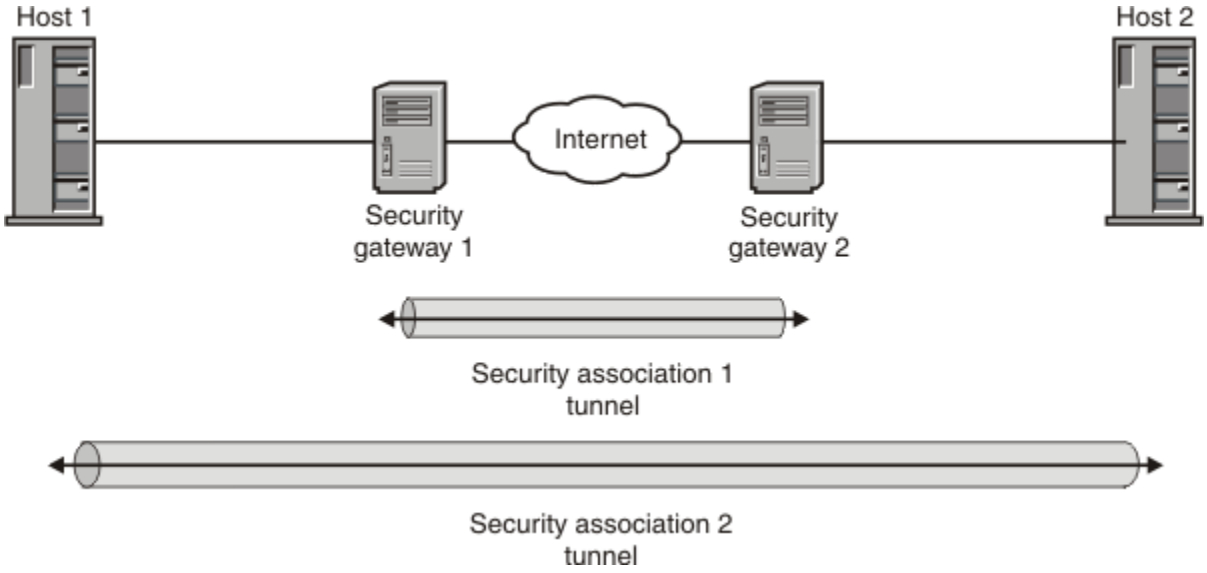


Figure 131. Nested tunnels

Mobile users

Mobile users represent a unique security model, because their IP address is not fixed and can be unpredictable. Instead of selecting traffic based on remote IP address, the mobile user's IKE identity can be used to select traffic. IPSec protection is required so that the IKE identity is known. You configure this security model similarly to other security models that require IPSec protection, except that the peer's remote identity is indicated on the RemoteIdentity parameter of the IpFilterRule statement, and the IpDestAddr parameter typically uses a wildcard value to indicate all addresses.

Multicast traffic

Multicast traffic can be protected by IPSec, but only manual tunnels are supported because the IKED supports negotiating dynamic tunnels with only a single peer rather than with a group of peers.

Multicast traffic is one-to-many (sent by individual nodes but received by multiple nodes) and is normally both sent and received; therefore, to use manual tunnels for multicast, you must use the same Security Parameter Index (SPI) and keys for inbound and outbound traffic. You must coordinate the SPI values and keys that are used with all multicast peers on the LAN segment. Also, because this manual tunnel is to be used to protect traffic with various source and destination addresses, you must specify any or any6 for the local and remote security endpoint locations. The following example shows AH authentication using the SHA algorithm, and ESP encryption using the DES algorithm.

```
IpManVpnAction tunnel-multicast
   {
     LocalSecurityEndpointAddr  any
     RemoteSecurityEndpointAddr any
     HowToAuth                  AH HMAC_SHA1
       AuthOutboundSa      2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
       AuthInboundSa       2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
     HowToEncrypt               DES
       EncryptOutboundSa   2701 0x3e6dcf72459ef551
       EncryptInboundSa    2701 0x3e6dcf72459ef551
     HowToEncap                 transport
   }
```

Requirement: You must define two filter rules for the multicast traffic. The first rule matches outbound multicast traffic, which has a unicast source address and a multicast destination address. The second rule matches inbound multicast traffic, which has a remote (destination) address that is unicast, and a local (source) address that is multicast. The addresses of the inbound rule are reversed from those that you might expect, because bidirectional rules are written from an outbound perspective. These rules are as follows:

```
IpFilterRule outbound-multicast
   {
     IpSourceAddrSetRef       lan-home-address
     IpDestAddr               224.0.0.1
     IpServiceRef             service-udp
     IpGenericFilterActionRef ipsec-nolog
     IpManVpnActionRef        tunnel-multicast
   }
   IpFilterRule inbound-multicast
   {
     IpSourceAddr             224.0.0.1
     IpDestAddr               lan-subnet
     IpServiceRef             service-udp
     IpGenericFilterActionRef ipsec-nolog
     IpManVpnActionRef        tunnel-multicast
   }
```

It is possible to restrict the tunnel to the multicast address that is being used. Define separate tunnels for different multicast addresses and use the same SPI value (the combination of address and SPI makes the tunnel unique). Because the local system is expected to participate in both sending and receiving multicast messages, you must create two manual tunnels. The following example shows this approach. In this example, one endpoint address is known for each tunnel, so you specify that address for the particular security endpoint address.

```
IpManVpnAction tunnel-multicast-outbound
   {
     LocalSecurityEndpointAddr  any
     RemoteSecurityEndpointAddr 224.0.0.1
     HowToAuth                  AH HMAC_SHA1
       AuthOutboundSa      2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
       AuthInboundSa       2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
     HowToEncrypt               DES
       EncryptOutboundSa   2701 0x3e6dcf72459ef551
       EncryptInboundSa    2701 0x3e6dcf72459ef551
     HowToEncap                 transport
   }
   IpManVpnAction tunnel-multicast-inbound
   {
     LocalSecurityEndpointAddr  224.0.0.1
     RemoteSecurityEndpointAddr any
     HowToAuth                  AH HMAC_SHA1
       AuthOutboundSa      2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
       AuthInboundSa       2700 0xa66e1b72e58a367ebd39d62daef84d5d9222cfe1
     HowToEncrypt               DES
       EncryptOutboundSa   2701 0x3e6dcf72459ef551
       EncryptInboundSa    2701 0x3e6dcf72459ef551
     HowToEncap                 transport
   }
   IpFilterRule outbound-multicast
   {
     IpSourceAddrSetRef       lan-home-address
     IpDestAddr               224.0.0.1
     IpServiceRef             service-udp
     IpGenericFilterActionRef ipsec-nolog
     IpManVpnActionRef        tunnel-multicast-outbound
   }
   IpFilterRule inbound-multicast
   {
     IpSourceAddr             224.0.0.1
     IpDestAddr               lan-subnet
     IpServiceRef             service-udp
     IpGenericFilterActionRef ipsec-nolog
     IpManVpnActionRef        tunnel-multicast-inbound
   }
```

Tip: Configuration of manual tunnels for IPv6 multicast is similar. For specific examples of configuring this for OSPFv3 security, see “Considerations for IPv6 OSPF security” on page 899.

Configuration scenarios supported for NAT traversal

Communications Server can act as a host Security Association endpoint for UDP-encapsulated mode Security Associations that are negotiated to enable IPSec traffic to traverse a NAT. The partner company with NAT model and the partner company with NAPT model describe Communications Server's host-to-

host support. The branch office with NAT model describes Communications Server's host-to-security gateway support.

Rule: Communications Server does not support acting as a security gateway endpoint for UDP-encapsulated tunnel mode Security Associations. This is different than the Communications Server support provided for tunnel mode Security Associations, where Communications Server can act as a security gateway, although Communications Server is not typically deployed in this manner.

The figures in the subtopics show z/OS configuration support for UDP-encapsulated Security Associations. A Security Association is negotiated by two IKE peers, with one initiating the negotiation and the other acting in responder mode. The location of the NAT, and the NAT's functionality, affect which IKE peer can initiate the Security Association. A traditional dynamic NAT implementation requires outbound traffic to be sent first to create an address mapping, before inbound traffic can be accepted. The dynamic NAT can be creating one-to-one address mappings from a dynamic pool of public IP addresses, or creating many-to-one address port mappings using a single public IP address and a pool of port values. When an IKE responder is behind a NAT, the NAT's address mapping must be static, allowing inbound traffic for the address to be received prior to outbound traffic being sent.

Host-to-host scenario 1 - z/OS-to-z/OS

Figure 132 on page 1037 shows a NAT in front of both z/OS hosts. A configuration with a NAT in front of only one of the z/OS hosts is supported as well. If there is a NAT device in front of the responder, the NAT's address mapping must be static. If there is a NAT device in front of the initiator, the NAT's address mapping can be static or dynamic. A dynamic mapping can use either one-to-one address translation or many-to-one address port translation (NAPT).

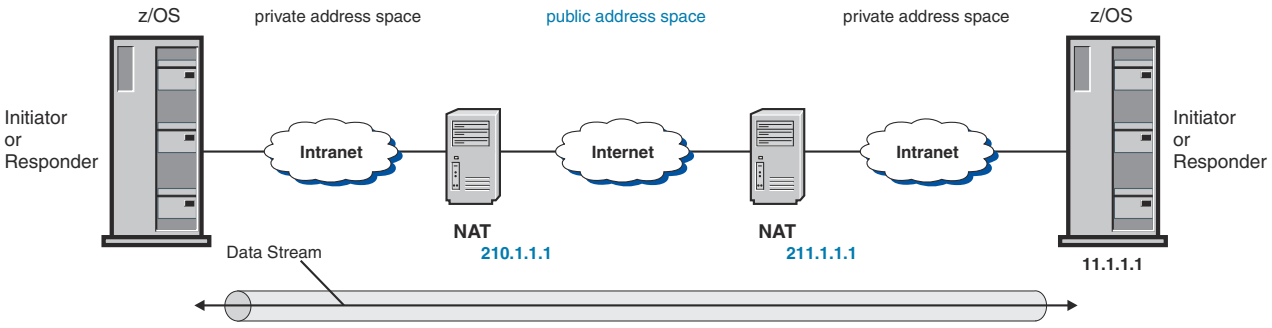


Figure 132. z/OS host to z/OS host, double NAT

Either UDP-encapsulated transport mode or UDP-encapsulated tunnel mode can be negotiated in a z/OS host-to-z/OS host configuration.

Rule: The z/OS host is limited to acting in responder mode when the remote endpoint is behind a NAPT. The negotiation of the phase 1 and phase 2 Security Associations must be initiated by the client behind the NAPT. Data must be initiated by the client behind the NAPT.

Host-to-host scenario 2 - z/OS-to-non-z/OS

Figure 133 on page 1038 shows a NAT in front of the z/OS host and the non-z/OS host. A configuration with a NAT in front of only one of the hosts is supported as well. If there is a NAT device in front of the responder, the NAT's address mapping must be static. If there is a NAT device in front of the initiator, the NAT's address mapping can be static or dynamic. A dynamic mapping can use either one-to-one address translation or many-to-one address port translation (NAPT).

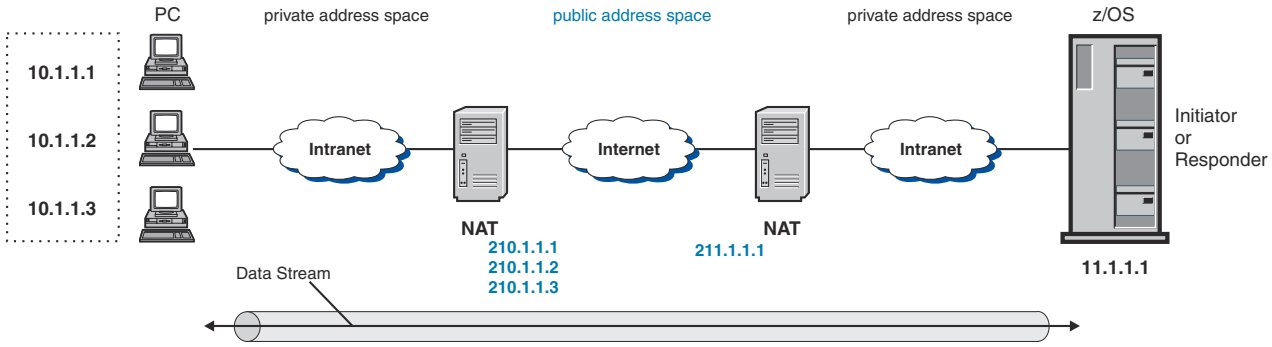


Figure 133. z/OS host to non-z/OS host, double NAT

Either UDP-encapsulated transport mode or UDP-encapsulated tunnel mode can be negotiated in a z/OS host-to-non-z/OS host configuration.

Rule: The z/OS host is limited to acting in responder mode when the remote endpoint is behind a NAPT. The negotiation of the phase 1 and phase 2 Security Associations must be initiated by the client behind the NAPT. Data must be initiated by the client behind the NAPT.

Interoperability Considerations

z/OS is typically used to provide a server function. The client initiates the phase 2 Security Association (SA) and data, with z/OS acting in the role of IKE responder and data responder. z/OS provides robust NAT traversal responder support, allowing it to interoperate with a variety of clients.

z/OS can also act as the initiator of the phase 2 SA and data. Potential incompatibilities exist in the following cases, depending on the support that the non-z/OS peer provides:

- Any IKEv2 tunnel mode phase 2 SA

  When z/OS initiates an IKEv2 tunnel mode phase 2 SA, it represents the data that is being protected by the SA with the following addresses:

  – The local IP address as it appears in the home list. This address might be a private IP address if z/OS is behind a NAT.

  – The public IP address of the client.

  The phase 2 SA negotiation should succeed if the non-z/OS peer supports receiving a tunnel mode traffic specification that is using these IP addresses.
- IKEv1 phase 2 SAs that protect specific ports, protocols, or both

  When initiating this type of phase 2 SA, z/OS represents the data being protected by the SA with the following addresses and values:

  – The local IP address as it appears in the home list. This could be a private IP address if z/OS is behind a NAT.

  – The client's public IP address.

  – The specific port and protocol values.

  The Phase 2 SA negotiation should succeed if the non-z/OS peer supports receiving this specification.
- IKEv1 tunnel mode phase 2 SAs that protect all ports and protocols

  When z/OS initiates this type of phase 2 SA, it does not explicitly include the IP addresses of the data being protected in the negotiation of the SA. This allows the non-z/OS peer to view the protected data in terms of the IP addresses that it understands, the public address of the remote endpoint and the private address of the local endpoint. The phase 2 SA negotiation should be successful.

  If data is initiated from z/OS over the SA, the data packet contains the following addresses:

  – The local IP address as it appears in the home list. This address might be a private IP address if z/OS is behind a NAT.

– The client's public IP address.

If the non-z/OS peer supports receiving a packet with these IP addresses, the data flow should be successful. When z/OS receives data packets from the non-z/OS peer, z/OS sends packets that contain the IP addresses used by the peer.

Host-to-security gateway scenario

Figure 134 on page 1039 shows both the security gateway and the host behind a NAT. z/OS also supports acting in responder mode when only one endpoint (either the security gateway or the z/OS host) is behind a NAT. When there is a NAT device in front of the z/OS host (acting as responder), the address mapping of the NAT must be static. If there is a NAT device in front of the security gateway, the address mapping of the NAT can be static or dynamic. A dynamic mapping can use either one-to-one address translation or many-to-one address port translation (NAPT).

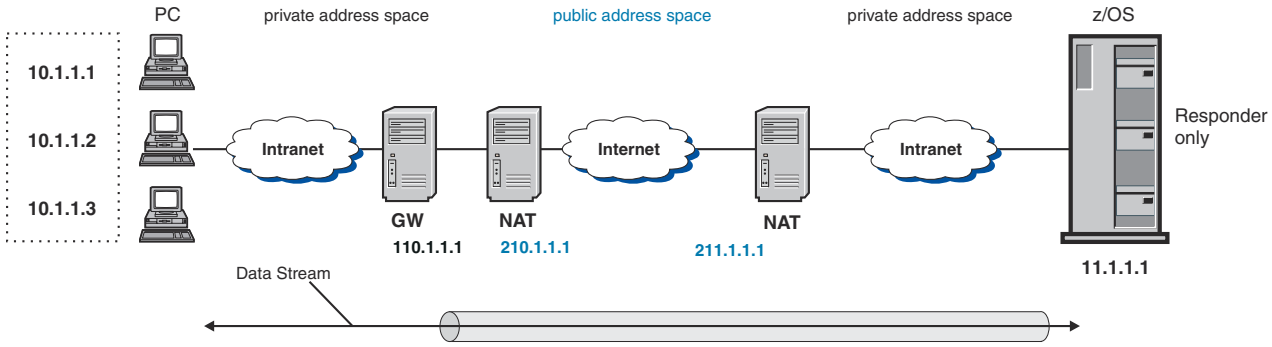


Figure 134. z/OS in a host-to-security gateway configuration

Rule: The z/OS host is limited to acting in responder mode in a host-to-security gateway configuration when a NAT is traversed. The phase 2 Security Association negotiation must be initiated by the security gateway. Data must be initiated by the client behind the security gateway.

Figure 134 on page 1039 shows the clients and the security gateway as separate devices. However, whenever a Security Association is negotiated to protect something other than a single IP address (for example, a range of IP addresses), that IKE daemon negotiating the Security Association is acting as a security gateway.

When the security gateway is behind a NAT, the individual hosts behind the NAT cannot be distinguished. If only one NAT address is available, all Security Associations negotiated between the security gateway (GW) and z/OS are negotiated using the NAT address, and have the same security characteristics. If multiple security characteristics are required to protect the traffic behind the security gateway, more NAT addresses are needed so that z/OS can locate different policies based on the NAT address.

Considerations for IPSec-encapsulated FTP traffic when traversing a NAT

FTP requires both a control connection and a data connection. For active-mode FTP, the client initiates the control connection and provides IP address and port information for the server to initiate the data connection. If the client is behind a NAT, the client provides its private IP address. The NAT updates the IP address in the FTP data for a packet that is not encapsulated; however, when you use IPSec encapsulation to secure your FTP connection, the NAT is unable to update the IP address in the FTP data.

Rule: When the FTP client is behind a NAT, you must use passive-mode FTP.

For passive-mode FTP, the client initiates the control connection. When a file transfer is to be started, the client sends a passive (PASV) command to the server. In response to the PASV command, the server provides the IP address and port information for establishing the data connection. The client is then able to initiate the data connection to the IP address and port.

When the FTP server is behind a NAT, the IP address provided in the PASV response is its private address. The NAT updates the IP address in the FTP data for a packet that is not encapsulated; however, when you use IPSec encapsulation to secure your FTP connection, the NAT is unable to update the IP address in the FTP data.

Rule: When the FTP server is behind a NAT, you must use extended passive-mode FTP.

Extended passive-mode FTP eliminates the IP address from the FTP data. The server provides the port for the data connection, and the client connects to the same IP address that was used for the control connection.

The z/OS FTP client and server support extended passive-mode FTP using the EPSV command. If you are using the z/OS FTP client and the FTP server that you are accessing does not support the EPSV command, you can configure the PASSIVEIGNOREADDR statement for the z/OS FTP client and use passive-mode FTP, which is widely supported. The PASSIVEIGNOREADDR statement directs the z/OS FTP client to ignore the IP address in the PASV reply, and to connect to the same IP address used for the control connection.

For more information about FTP, see Chapter 12, "Transferring files using FTP," on page 663. For a sample FTP.DATA data set, or for information about the LOCSIte subcommand, see z/OS Communications Server: IP User's Guide and Commands.

Enterprise Extender considerations when traversing a NAT

The implementation of Enterprise Extender (EE) requires that the EE connection endpoints be defined by unique static VIPA addresses. NAT functions are limited in the EE environment as follows:

- The NAT mapping must be a one-to-one address mapping. NAPT is not supported.
- Dynamic mappings are generally unreliable for an EE connection. A static mapping of internal IP address to external IP address should be defined when an EE endpoint is behind a NAT.
- When IPSec protection is added for EE traffic that traverses a NAT, only one host that is behind a security gateway that is behind a NAT will be able to send EE traffic. In most cases, EE hosts should not be located behind a security gateway that is behind a NAT. Instead, a host-to-host Security Association should be negotiated for each EE host.

Additional configuration concerns for NAT traversal

The following list contains some additional configuration concerns for NAT traversal:

- When using NAT traversal, z/OS views its own address as the one configured in the z/OS home list. If the z/OS host is behind a NAT, this address is a private address. Otherwise, it is the public address of the z/OS host.

  The z/OS implementation does not use private addresses in its configuration to describe the remote IKE peer or remote IP connection endpoint. z/OS views its IPSec peer and the remote IP connection endpoint as a public IP address. If a NAT is in front of the IPSec peer, the z/OS host perceives the IPSec peer and connection endpoint addresses to be that of the NAT.
- For IKEv1 Security Associations, you should not configure pre-shared keys in Main mode when multiple remote peers are behind a NAT and the peers do not map to unique RemoteSecurityEndpoint specifications. RFC 3947 states that pre-shared keys cannot be used with Main mode, unless group shared keys for all those behind the NAT are deployed. The use of group pre-shared keys is considered a security risk.
- If a NAT address is coming out of a dynamic NAT pool, addresses assigned to a host from the pool can be assigned any of the pooled addresses. In this case, there are additional considerations. When z/OS is the responder, the IPSec policy intended for a host must be broad enough to cover the entire range of IP addresses in the dynamic NAT pool. When z/OS is the initiator and the responder is behind a NAT, the identity of the target host can be ambiguous. Do not configure a z/OS host to initiate to an ambiguous target host.
- When a remote security endpoint is behind a NAT, its identity must be unique. During a phase 1 negotiation, the remote security endpoint sends its identity in an ID payload. The IKE daemon can manage multiple remote security endpoints using the same ID when those endpoints are not behind a NAT. However, when a remote security endpoint is behind a NAT, it must use a unique IKE identity.
- When the remote security endpoint is a security gateway behind a NAT or a host behind a NAPT, only TCP, UDP, and ICMP traffic is supported. ICMP traffic has limited support.

- z/OS allows traffic for a TCP connection to continue as long as the integrity of the connection can be verified. Two cases where z/OS can no longer verify the integrity of the connection are:
  - Adding or removing IPSec protection for a TCP connection

    When a TCP connection traverses a NAT, the TCP connection must be restarted after a filter policy change that causes the connection's traffic to change from IPSec-protected traffic to clear text, or from clear text to IPSec-protected traffic.
  - NAT IP address remapping

    If the peer's IP address is remapped by a NAT due to a timeout or reboot of the NAT device, the TCP connection must be restarted.

Configuring the IKE daemon

This topic describes considerations and steps for configuring the IKE daemon.

The IKE daemon's purpose is to manage dynamic IPSec tunnels and to provide a network management interface (NMI) for monitoring and controlling IP filtering and IPSec. The IKE daemon is not involved in the actual filtering, encapsulation, or decapsulation of packets. The IKE daemon is not required for the configuration or use of IP filters when no IpDynVpnAction statements are used. However, because the IKE daemon processes NMI monitoring requests, it must be running to gather monitoring data for IP filters, manual Security Associations, or dynamic Security Associations. To start the IKE daemon, it must be able to connect to the Policy Agent. For information about this requirement, see "Policy Agent considerations" on page 1042. For more information about the IPSec network management interface, see z/OS Communications Server: IP Programmer's Guide and Reference.

Multiple TCP/IP stacks

A one-to-many relationship can exist between an instance of the IKE daemon and stacks configured with IPCONFIG IPSECURITY. A single instance of the IKE daemon can service all stacks configured with IPCONFIG IPSECURITY on a single z/OS image. Only one instance of the IKE daemon can run on a single z/OS image.

Each stack can be configured as a network security services (NSS) client. An NSS client makes use of network security services offered by the NSS server. For details about configuring an NSS server, see Chapter 18, "Network security services," on page 1083.

"TCP/IP stack initialization access control" on page 146 describes a time interval during which limited stack access is available for stacks that have been configured for AT-TLS using the TCPCONFIG statement with the TTLS parameter. To enable the IKE daemon for a stack during this interval, the IKE daemon user ID must be permitted to the EZB.INITSTACK.*sysname.tcpname* resource profile. For examples of the security product commands needed to grant access to this profile, see member EZARACF in sample data set SEZAINST.

Run-time environment

The IKE daemon is a z/OS UNIX application, and it requires a z/OS UNIX file system such as z/OS File System. The IKE daemon can be started from an MVS started procedure, from the z/OS shell, with the AUTOLOG statement in the TCP/IP profile, or by using the COMMNDxx member of PARMLIB. The IKE daemon must be started by a RACF-authorized user ID, and it must be in an APF-authorized library. For more information about how to start the IKE daemon, see "Starting the IKE daemon" on page 1048.

The IKE daemon uses the MVS operator's console, syslogd, CTRACE, and STDOUT for its logging and tracing. The MVS operator's console and STDOUT are used for major events such as initialization, termination, and error conditions. Syslogd is used for logging events related to dynamic IPSec tunnel management. CTRACE is used for detailed tracing and debugging.

The IKE daemon uses a standard message catalog. The message catalog must be in the z/OS UNIX file system. The directory location for the message catalog path is set by the environment variables NLSPATH and LANG.

Language Environment run-time considerations

When starting the IKE daemon from a started or cataloged procedure, you should usually start the IKE daemon directly from the SEZALOAD data set using PGM=IKED. However, there is a situation where you might want to start the IKE daemon using BPXBATCH.

When the IKE daemon is started using PGM=IKED, the STDENV DD card, if used, is passed directly to the IKE daemon program. Language Environment does not get access to the STDENV environment variables. As a result, any Language Environment run-time options set in the STDENV DD data set using the _CEE_RUNOPTS= environment variable are ignored. In this case, Language Environment run-time options must be passed on the PARM= parameter and the options must be specified before any IKE daemon options. However, the PARM= statement allows a maximum of 100 characters. If the wanted Language Environment run-time options plus IKE daemon parameters exceeds 100 characters, consider using BPXBATCH to start the IKE daemon. When PGM=BPXBATCH is used, the Language Environment variable _CEE_RUNOPTS can be included on the STDENV DD card to specify run-time options in excess of 100 characters long.

IKE daemon configuration source information

The IKE daemon obtains configuration information from two sources.

- IKE daemon configuration file

  The IKE daemon configuration file contains the IkeConfig statement. Parameters on the IkeConfig statement are global IKE daemon operational parameters. For details on the IKE daemon configuration file and the IkeConfig statement, see z/OS Communications Server: IP Configuration Reference.

  The IKE daemon configuration file is read when the IKE daemon initializes and can be reread dynamically. For more information, see “Controlling the IKE daemon” on page 1050.

- Policy Agent

  IP security policy includes dynamic IPSec tunnel configuration information required by the IKE daemon. The IKE daemon obtains IP security policy from the Policy Agent. The IKE daemon obtains IP security policy when connecting to the Policy Agent and whenever the Policy Agent informs the IKE daemon of a change in IP security policy. The IKE daemon connects only to stacks configured with IPCONFIG IPSECURITY, if there is an IP security policy defined for a stack. For details on configuring IP security policy, see “Configuring specific security models” on page 974.

Policy Agent considerations

The IKE daemon cannot perform management of dynamic IPSec tunnels until it has obtained IP security policy from the Policy Agent. While the Policy Agent is running, it can inform the IKE daemon of dynamic changes to IP security policy. After the IKE daemon has obtained an IP security policy, the Policy Agent can be stopped without impacting the IKE daemon. However, any changes to the IP security policy will not be detected until the Policy Agent is restarted, nor will IKE detect newly activated or reactivated stacks. The IKE daemon reconnects to the Policy Agent when the Policy Agent is restarted. For information about starting and monitoring policy related applications using the Policy Agent, see “Step 7: Configuring Policy Agent to automatically monitor applications” on page 820.

Using network security services

The IKE daemon allows a stack to be defined as a network security services (NSS) client. When a stack is defined as an NSS client, the IKE daemon uses at least one network security service on behalf of that stack. Network security services are provided by an NSS server. An NSS server provides a certificate service and a remote management service. For details about the configuration of an NSS server, see Chapter 18, “Network security services,” on page 1083.

The certificate service of the NSS server is used to create and verify digital signatures on behalf of an NSS client. Certificates for stacks that are configured to use the NSS certificate service must be on the key ring of the NSS server. For details about configuring the IKE daemon to use the NSS certificate service, see “Step 6: Setting up the IKE daemon for digital signature authentication (optional)” on page 1367.

Restriction: If you want the IKED to use a digital signature authentication method to negotiate an IKEv2 Security Association for a stack, the stack must be configured to use the NSS certificate service.

The remote management service of the NSS server enables the IP filter rules and Security Associations of an NSS client to be monitored and managed from the system on which the NSS server is executing. For details about using the **ipsec** command to monitor and manage NSS clients, see z/OS Communications Server: IP System Administrator's Commands. For details about using the IPSec NMI to monitor and manage NSS clients, see z/OS Communications Server: IP Programmer's Guide and Reference.

The NSS server does not need to be on the same system as the IKE daemon. The location of the NSS server is specified by the NetworkSecurityServer parameter and optionally by the NetworkSecurityServerBackup parameter of the IkeConfig statement. For additional details about the IkeConfig statement, see z/OS Communications Server: IP Configuration Reference

The NssStackConfig statement in the IKE daemon configuration file is used to define a stack as an NSS client. The ServiceType parameter of the NssStackConfig statement identifies what network security services are to be used by a stack. When a ServiceType value Cert is specified, the NSS certificate service is used. When a ServiceType value RemoteMgmt is specified, the NSS remote management service is used. The ServiceType parameter can be specified multiple times. For complete details about the NssStackConfig statement, see z/OS Communications Server: IP Configuration Reference.

When a stack is configured to use the NSS remote management service, the NSS server can send the IKE daemon a request to switch between default IP filter policy and IP security filter policy. To change filter sets, the IKE daemon creates or deletes a specific marker file that the stack accesses. This marker file is the same marker file that is created or deleted by the **ipsec** command when the **ipsec -f reload** or **ipsec -f default** IP filter options are specified locally without the -z option. For details about the marker file, see z/OS Communications Server: IP System Administrator's Commands.

The ClientName parameter of the NssStackConfig statement associates an NSS client name with a stack. This is the name by which the NSS server knows the stack. The UserId parameter of the NssStackConfig statement associates the NSS client name with a user ID defined on the NSS server's system. The NSS server uses both the client name and user ID when checking SERVAUTH profiles to verify that an NSS client is authorized to request a specific action. For details about SERVAUTH profiles checked by the NSS server, see Chapter 18, “Network security services,” on page 1083.

The AuthBy parameter of the NssStackConfig statement defines how the user ID associated with a stack that is acting as an NSS client is to be authenticated. This user ID can be authenticated using a password or a PassTicket. When a PassTicket is used, the application key is stored in the local external security manager database.

Tip: Using a PassTicket is more secure than specifying a password.

To store the application key in the local external security manager database, the secure signon function must be enabled and a PTKTDATA profile must be created. This key must be associated with an application ID of the NSSD. For specific information about enabling the secure signon function and defining profiles to be used by the single signon function, see z/OS Security Server RACF Security Administrator's Guide.

The following command is an example of a RACF command that you can issue to store the application key for the NSS server and NSS clients:

```
RDEFINE PTKTDATA NSSD SSIGNON(KEYMASKED(E001193519561977)) UACC(NONE)
```

Figure 135 on page 1044 shows a partial configuration for the IKE daemon on system SYSTEMA. The NetworkSecurityServer parameter on the IkeConfig statement specifies that the IKE daemon is configured to use network security services from an NSS server that is listening on IP address 9.1.1.1. One NssStackConfig statement is shown. The ClientName parameter associates stack STACK1 with an NSS client name SYSTEMA_STACK1. This is the name by which the NSS server knows this stack. The UserId parameter associates the client name with the user ID A1S1. The A1S1 user ID must be defined on the NSS server's system. Both the client name and user ID are used by the NSS server when verifying the authorization of an NSS client. The multiple ServiceType parameters indicate that the IKE daemon uses both the NSS certificate service and the NSS remote management service.

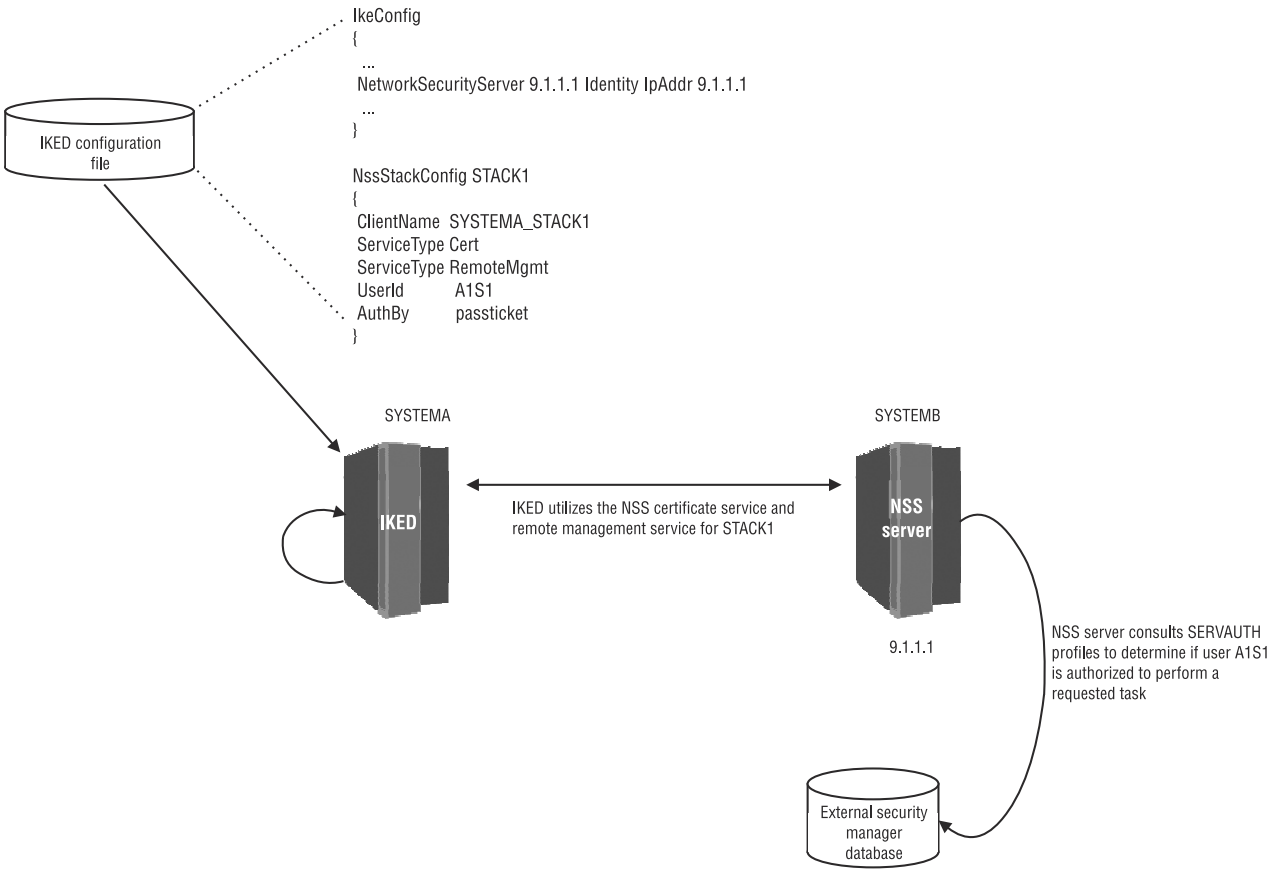


Figure 135. Enabling network security services

The IKE daemon requires that communication with the NSS server be protected using AT-TLS. During the AT-TLS handshake, the NSS server provides a certificate that can be used to authenticate its identity. The IKE daemon examines this certificate and verifies that the identity in the certificate matches the identity specified on the NetworkSecurityServer parameter of the IkeConfig statement.

The IKE daemon does not perform any SERVAUTH checks when processing an IPSec monitoring request or an IPSec management request from the NSS server. The NSS server performs SERVAUTH checks to make sure that the requester of an IPSec monitoring or management request is authorized. For details about the SERVAUTH requirements imposed by the NSS server, see z/OS Communications Server: IP Programmer's Guide and Reference.

The IKE daemon does perform a SERVAUTH check for the EZB.NETMGMT.*sysname*.*sysname*.IKED.DISPLAY profile when processing a local NMI request to display information about current NSS client state. For additional details, see z/OS Communications Server: IP Programmer's Guide and Reference.

Certificate revocation checking

Certificate revocation checking is applicable only to digital signature authentication methods. The RevocationChecking parameter in the IPSec policy file controls the level of certificate revocation checking that is performed during an IKE negotiation. The following three levels of revocation checking are supported:

- Strict
- Loose
- None

You can specify the RevocationChecking parameter on the KeyExchangePolicy statement and the KeyExchangeAction statement. For more information about the KeyExchangePolicy statement and the KeyExchangeAction statement, see z/OS Communications Server: IP Configuration Reference.

The native IKED certificate service does not support the retrieval and checking of certificate revocation information. When the IKED is configured to use the native IKE daemon certificate service, the RevocationChecking parameter is ignored.

The NSS certificate service does support the retrieval and checking of certificate revocation information in the form of certificate revocation lists (CRLs). For information about the NSS server requirements for retrieving CRLs, see "NSS server certificate revocation support" on page 1104. Ensure that these requirements can be met before you enable strict revocation checking in the IPSec policy file.

Steps for configuring the IKE daemon

The IKE daemon manages dynamic IPSec tunnels and provides a network management interface (NMI) for monitoring and controlling IP filtering and IPSec. Because the IKE daemon processes NMI monitoring requests, it must be running to gather monitoring data for IP filters, manual Security Associations, or dynamic Security Associations.

Procedure

Perform the following steps to configure the IKE daemon:

1. Create the IKE daemon configuration file.

   A sample configuration file is provided in /usr/lpp/tcpip/samples/iked.conf.

   The following search order is used by the IKE daemon to locate the configuration data set or file:

   a. If the environment variable IKED_FILE has been defined, the IKE daemon uses the value as the name of an MVS data set or z/OS UNIX file to access the configuration data.

   b. /etc/security/iked.conf

   You can specify statements in the configuration file using a variety of EBCDIC code pages. Use the IKED_CODEPAGE environment variable to specify the code page that you want to use. The default code page is IBM-1047.

2. Set the _BPX_JOBNAME environment variable (optional).

   When starting the IKE daemon from the z/OS shell, the environment variable _BPX_JOBNAME should be set. This enables a specific job name to be used when reserving ports for the IKE daemon. This name can also be used with the STOP or MODIFY console commands.

   For more information on _BPX_JOBNAME, see z/OS UNIX System Services Planning

3. Reserve the ports.

   Update the PORT statement in PROFILE.TCPIP to reserve ports 500 and 4500 for the IKE daemon. Add the name of the member containing the IKE daemon cataloged procedure or the name as set using _BPX_JOBNAME:

   ```
   PORT
                    500 UDP IKED
                    4500 UDP IKED
   ```

4. Update the IKE daemon cataloged procedure.

   If the IKE daemon is to be started by a procedure, create the cataloged procedure by copying the following sample in SEZAINST(IKED) to your system or recognized PROCLIB. Specify IKE daemon parameters and change the data set names to suit your local configuration.

```
//IKED     PROC
//*
//* IBM Communications Server for z/OS
//* SMP/E distribution name: EZBIKPRC
//*
```

```
//* 5650-ZOS Copyright IBM Corp. 2005, 2013
//* Licensed Materials - Property of IBM
//* "Restricted Materials of IBM"
//* Status = CSV2R1
//*
//*
//IKED     EXEC PGM=IKED,REGION=0K,TIME=NOLIMIT,
//       PARM='ENVAR("_CEE_ENVFILE_S=DD:STDENV")/'
//*
//* Provide environment variables to run with the desired
//* configuration.  As an example, the data set or file specified by
//* STDENV could contain:
//*
//*   IKED_FILE=/etc/security/iked.conf2
//*   IKED_CTRACE_MEMBER=CTIIKE01
//*   IKED_CODEPAGE=IBM-1047
//*
//*
//*   If you want to include comments in the data set or
//*   z/OS UNIX file, specify the _CEE_ENVFILE_COMMENT
//*   environment variable as the first environment variable
//*   in the data set or file.  The value specified for
//*   the _CEE_ENVFILE_COMMENT variable is the comment character.
//*   For example, if you want to use the pound sign, #, as
//*   the comment character, specify this as the first
//*   statement:
//*     _CEE_ENVFILE_COMMENT=#
//*
//* For information on the above environment variables, refer to the
//* IP Configuration Reference.
//*
//STDENV   DD DUMMY
//* Sample MVS data set containing environment variables:
//*STDENV   DD DSN=TCPIP.IKED.ENV(IKED),DISP=SHR
//* Sample HFS file containing environment variables:
//*STDENV   DD PATH='/etc/security/iked.env',PATHOPTS=(ORDONLY)
//*
//* Output written to stdout and stderr goes to the data set or
//* file specified with SYSPRINT or SYSOUT, respectively.
//SYSPRINT DD SYSOUT=*
//SYSOUT   DD SYSOUT=*
```

5. Authorize the IKE daemon to the external security manager.

   See “Step 3: Authorizing the IKE daemon to the external security manager” on page 1362.

6. Configure and start syslogd.

   The IKE daemon uses the local4 facility when writing messages to syslogd. For performance purposes, syslogd should use z/OS File System as its underlying file system. For more information on syslogd, see “Configuring the syslog daemon” on page 213.

   Tip: The system logging daemon (syslogd) can be configured to forward messages from the IKE daemon to a syslogd on another host. For information about forwarding syslog messages to another host, see z/OS Communications Server: IP Configuration Reference. When a stack is configured as an NSS client, it can be advantageous to forward syslog messages from the IKE daemon to the syslogd running on the NSS server's system. Configuring syslogd in this manner allows all IKE messages relating to an NSS client to be in the same log file as the NSS server's messages.

7. Update the IKE daemon environment variables (optional).

   The following environment variables are used by the IKE daemon and can be tailored to a particular installation:

   IKED_CODEPAGE
   Use the IKED_CODEPAGE variable to specify the EBCDIC code page to be used when reading the configuration file. For more information about IKE environment variables and the supported code pages, see z/OS Communications Server: IP Configuration Reference.

IKED_CTRACE_MEMBER
The IKED_CTRACE_MEMBER variable is used by the IKE daemon to locate a parmlib member for IKE daemon CTRACE customization. For more information on the TCP/IP services component trace for the IKE daemon, see z/OS Communications Server: IP Diagnosis Guide.

IKED_FILE
The IKED_FILE variable is used by the IKE daemon in the search order for the IKE daemon configuration file. For details on the search order used for locating this configuration file, see step "1" on page 1045.

8. Setup the IKE daemon for TCP/IP stack initialization access control (optional).

   See "Multiple TCP/IP stacks" on page 1041.

9. Setup the IKE daemon for digital signature mode authentication (optional).

   See "Step 6: Setting up the IKE daemon for digital signature authentication (optional)" on page 1367.

10. Define AT-TLS policy to protect communication with an NSS server.

    The IKE daemon requires that communication between the NSS server and the IKE daemon be secured using Application Transparent Transport Layer Security (AT-TLS). If a stack is configured as an NSS client, AT-TLS rules must be defined to secure this communication. Enable AT-TLS processing for a stack by specifying the TTLS parameter on the TCPCONFIG statement in the TCP/IP profile. Specific AT-TLS policy is configured in Policy Agent configuration files. For details about enabling AT-TLS and configuring AT-TLS policy, see Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121.

    Tip: Define AT-TLS policy such that only cipher suites requiring TLS encryption are exchanged with the NSS server. Failure to restrict the cipher suites to those requiring encryption can result in sensitive information flowing in the clear across an untrusted network.

    Rule: AT-TLS policy must be defined for each stack through which the IKE daemon communicates with the NSS server.

    A sample AT-TLS policy is located in /usr/lpp/tcpip/samples/pagent_TTLS.conf.

    Rule: The RemotePortRange value in the TTLSRule statement must include the value specified on the NetworkSecurityServer port parameter or the NetworkSecurityServerBackup port parameter in the IKE daemon configuration file.

11. Define IP filter policy to enable communication with an NSS server (optional).

    If a stack is configured as an NSS client, IP filter policy for that stack must be defined to enable this communication. The IKE daemon communicates with the NSS clients using the TCP protocol. By default, the NSS server listens on port 4159. The IKE daemon connects to the NSS client using an ephemeral port. Ephemeral ports are generally in the range 1024 – 65355.

    Two types of IP filter policy can be defined for a z/OS stack:

    - Default IP filter policy is defined in the TCP/IP profile. Updating default IP filter policy to permit communications between the IKE daemon and the NSS server is optional. Default IP filter policy is in effect only when IP security filter policy cannot be loaded or when the **ipsec -f default** command has been issued. For details about how to define default IP filter policy, see z/OS Communications Server: IP Configuration Reference.

      The following example of a default policy contains IPSECRule definitions that allow IKE daemon traffic with the NSS server:

```
IPSEC LOGENable
; Rule      SrcAddr DstAddr   Logging Protocol   SrcPort    DestPort     Routing Secclass

; OSPF protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO OSPF

; IGMP protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO 2

; DNS queries to UDP port 53
  IPSECRule *       *         NOLOG   PROTO UDP  SRCPort *  DESTport 53

; Administrative access
  IPSECRule *       9.1.1.2   LOG                                          SECCLASS 100
```

```
; IKE daemon access to the Network Security Server
  IPSECRule *       *         LOG     TCP        SRCPort *  DESTport 4159

; IKE daemon access to the Network Security Server
  IPSEC6Rule *      *         LOG     TCP        SRCPort *  DESTport 4159

ENDIPSEC
```

Rule: The DESTport value in the filter rules must include the value specified for the NetworkSecurityServer port parameter or the NetworkSecurityServerBackup port parameter in the IKE daemon configuration file.

- IP security filter policy is defined in Policy Agent configuration files. IP security filter policy must be updated to permit communications between the IKE daemon and the NSS server. For details about how to define IP security policy files, see z/OS Communications Server: IP Configuration Reference.

  The following example shows an IpFilterRule statement for IPv4, an IpFilterRule statement for IPv6, and an IpGenericFilterAction statement that allow the IKE daemon to communicate with the NSS server:

```
IpFilterRule             NssTrafficIPv4
{
    IpSourceAddr             all4
    IpDestAddr               all4
    IpService
    {
       SourcePortRange        1024 65535
       DestinationPortRange   4159
       Protocol               tcp
       Direction              bidirectional OutboundConnect
       Routing                local
    }
    IpGenericFilterActionRef  permit-nolog
}

IpFilterRule             NssTrafficIPv6
{
   IpSourceAddr             all6
   IpDestAddr               all6
   IpService
   {
      SourcePortRange        1024 65535
      DestinationPortRange   4159
      Protocol               tcp
      Direction              bidirectional InboundConnect
      Routing                local
   }
   IpGenericFilterActionRef  permit-nolog
}

IpGenericFilterAction     permit-nolog
{
   IpFilterAction           permit
   IpFilterLogging          no
}
```

Rule: The DestinationPortRange value on the IpService statement must include the value specified on the NetworkSecurityServer port parameter or the NetworkSecurityServerBackup port parameter in the IKE daemon configuration file.

Starting the IKE daemon

After the necessary external security manager authorization has been defined (see “Step 3: Authorizing the IKE daemon to the external security manager” on page 1362), the IKE daemon can be started from an MVS procedure, from the z/OS shell, or using the AUTOLOG statement.

- You can start the IKE daemon procedure from the MVS operator console. A sample start procedure is provided in SEZAINST(IKED).
- You can start the IKE daemon from the z/OS shell by starting OMVS and then issuing the iked command.

- You can use the AUTOLOG statement to start the IKE daemon automatically during TCP/IP initialization by inserting the name of the IKE daemon start procedure into the AUTOLOG statement in the PROFILE.TCPIP data set:

```
AUTOLOG
  IKED
ENDAUTOLOG
```

Tips:

  - When implementing multiple stacks enabled for IP security, adding an AUTOLOG statement for the IKE daemon might not be optimal. If the IKE daemon is listed in an AUTOLOG statement of a stack's profile, the IKE daemon is cancelled if it is already running when that stack starts. In a multiple IP security stack environment, this could disrupt traffic on other IP security stacks. Use another method to automate starting the IKE daemon when the system is IPLed, such as using the COMMNDxx member of PARMLIB. For more information about the use and configuration of the COMMNDxx member of PARMLIB, see z/OS MVS Initialization and Tuning Reference.
  - If you start the IKE daemon from the z/OS shell and you stop the shell environment from scrolling, then when the daemon needs to display data to the shell it might stop and wait indefinitely for the shell to scroll and make output buffer space available for the data.

When running from an MVS procedure, the environment variables can be set using the STDENV DD statement in the IKE daemon procedure. For information concerning the environment variables used by IKE daemon, see step "7" on page 1046 in "Steps for configuring the IKE daemon" on page 1045.

The /var/ike/iked.pid is a temporary IKE daemon pid file that the IKE daemon creates. This file contains the process ID of the current invocation of the IKE daemon.

Restrictions:

- If /var/ike/iked.pid is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the IKE daemon.
- If /var/ike/iked.pid is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /var/ike/iked.pid is stored cannot have write access to the directory. Additionally, write access to /var/ike/iked.pid must be limited to the owning UID or group, for example, --w--w----permissions.

Stopping the IKE daemon

The IKE daemon can be stopped as follows:

- From MVS, issue:

```
STOP procname
```

  If the IKE daemon was started from a cataloged procedure, *procname* is the member name of that procedure. If the IKE daemon was started from the z/OS shell and the environment variable _BPX_JOBNAME was set, *procname* is the same as _BPX_JOBNAME. If the IKE daemon was started from the z/OS shell and _BPX_JOBNAME was not set, *procname* is based on the *userid*. If the *userid* is 8-character long, the *procname* is the *userid*. If the *userid* is less than 8-character long, the *procname* is *useridX*, where X is the sequence number set by the system. To determine the sequence number, from the SDSF LOG window on TSO, issue:

```
/d omvs,u=userid
```

  This command shows the programs running under this user ID. For more information on _BPX_JOBNAME, see z/OS UNIX System Services Planning.
- From a superuser ID in the z/OS shell, issue the kill command to the process ID (PID) associated with the IKE daemon. The IKE daemon PID is recorded in /var/ike/iked.pid.

Controlling the IKE daemon

You can control the IKE daemon from the operator's console using the MODIFY command. MODIFY commands are available to perform the following functions:

- Rereading the configuration file

  The MODIFY *procname*,REFRESH command is used to reread the IKE daemon configuration file. Not all IkeConfig statement parameters can be updated using this command. For information on which parameters can be dynamically changed, see the parameter descriptions for the IkeConfig statement of the IKE daemon configuration file in the z/OS Communications Server: IP Configuration Reference.

- Displaying the configuration file parameters

  The MODIFY *procname*,DISPLAY command is used to display configuration values currently being used by the IKE daemon.

For more information on the MODIFY command, see z/OS Communications Server: IP System Administrator's Commands.

Verifying policy installation

This topic describes the console messages and commands that are used to verify policy installation.

Console messages

After the IP security policy has been configured, start the TCP/IP stacks, Policy Agent, and the IKED. A series of console messages is issued if the installation of IP security policy was successful.

The following console messages indicate that Policy Agent and IKE have completed initialization:

```
EZZ8432I PAGENT INITIALIZATION COMPLETE
EZD1046I IKE INITIALIZATION COMPLETE
```

The following console messages indicate that the processing of IP security policy is complete:

```
EZZ8771I PAGENT CONFIG POLICY PROCESSING COMPLETE FOR TCPCS : IPSEC
EZD1068I IKE POLICY UPDATED FOR STACK TCPCS
```

If there were errors in the configuration files, Policy Agent issues the following message to the console:

```
EZZ8438I PAGENT POLICY DEFINITIONS CONTAIN ERRORS FOR TCPCS : IPSEC
```

Look at the Policy Agent log to find and correct the error.

Displaying TCP/IP configuration

To display whether the TCP/IP stack is configured with IPCONFIG IPSECURITY, issue the **`netstat -f`** command and look for the following field in the `IPv4 Configuration Table` section:

```
IpSecurity: Yes
```

To display whether the TCP/IP stack is configured with IPCONFIG6 IPSECURITY, issue the **`netstat -f`** command and look for the following field in the `IPv6 Configuration Table` section:

```
IpSecurity: Yes
```

To display whether the TCP/IP stack is configured for Sysplex-Wide Security Associations, issue the **`ipsec -f display`** command. The DVIPSec field in the header of the command display shows whether or not the DVIPSEC keyword has been coded in the TCP/IP profile:

```
ipsec -f display | head -n 7
```

```
CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 10:52:03 2010
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Policy    Scope:    Current            TotAvail: 137
Logging:  On              Predecap: Off                DVIPSec:  Yes
NatKeepAlive:  20         FIPS140:  No
Defensive Mode: Inactive
```

Displaying active filters with the ipsec command

Use the **ipsec -f display** command to display active filter rules, configured filter rules from IP security policy configuration files, and the default IP filter rules from the TCP/IP profile. The scope on the command, as indicated by the -c option, determines which source is queried:

-c policy
Shows IP filters as configured in the IP security policy configuration files.

-c profile
Shows default IP filters as configured in the TCP/IP profile.

-c current
Shows active IP filters in the stack. The active filters that are shown can be the default IP filters as defined in the TCP/IP profile, or IP filters as configured in the IP security policy configuration files, depending on which policy is active at the time the command is issued. The output of the display indicates the source of the current active filters.

The output of the command can be quite voluminous, so you might want to redirect the output of the display to a file.

The information in the report header of the report output indicates how many filters are active, and also indicates the source of the filters, whether from the default IP filter policy or the IP security policy from the Policy Agent.

```
ipsec -f display

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 10:53:12 2010
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Profile   Scope:    Current            TotAvail: 14
Logging:  On              Predecap: Off                DVIPSec:  Yes
NatKeepAlive:  20         FIPS140:  No
Defensive Mode: Inactive
```

If the source field shows `Stack Policy`, the IP security policy is installed and active.

If the source field shows `Stack Profile`, the IP security policy is either not installed or the **ipsec -f default** command was issued. Either issue the **ipsec -f reload** command, or correct the IP security policy configuration.

Filter displays can be abbreviated to include only specific named rules. To view a named filter rule, use the -n option as follows:

```
ipsec -f display -n Rule2Admin

CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 10:54:36 2012
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Policy    Scope:    Current            TotAvail: 137
Logging:  On              Predecap: Off                DVIPSec:  Yes
NatKeepAlive:  20         FIPS140:  No
Defensive Mode: Inactive

FilterName:                   Rule2Admin
FilterNameExtension:          1
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
```

```
OnDemand:                       No
SecurityClass:                  0
Logging:                        Deny
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.1.1.1
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    9.1.1.2
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:49:48
UpdateTime:                     2012/02/14 10:49:48
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2Admin
FilterNameExtension:            2
GroupName:                      Admin
LocalStartActionName:           n/a
VpnActionName:                  Silver-TransportMode
TunnelID:                       Y0
Type:                           Dynamic Anchor
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Inbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        Deny
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.1.1.2
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    9.1.1.1
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
```

```
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:49:48
UpdateTime:                     2012/02/14 10:49:48
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*********************************************************************

2 entries selected
```

Anchor filters and dynamic filters

After a Security Association is negotiated, the **ipsec -f display** command shows the addition of two dynamic filters that were created when the Security Association was created, corresponding to the inbound and outbound anchor filters. Dynamic filters are placed ahead of the anchor filters in the filter table, so dynamic filters are searched first when IP filtering is performed. In the following sample output, note that two dynamic filters have been added to the filter table subsequent to the activation of a phase 2 Security Association. The Type field indicates whether the filter is a dynamic anchor filter or a dynamic filter:

```
ipsec -f dis -n Rule2Admin

CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 11:23:54 2012
Primary:  Filter          Function: Display            Format:   Detail
Source:   Stack Policy    Scope:    Current            TotAvail: 139
Logging:  On              Predecap: Off                DVIPSec:  Yes
NatKeepAlive:  20         FIPS140:  No
Defensive Mode: Inactive

FilterName:                     Rule2Admin
FilterNameExtension:            1
GroupName:                      Admin
LocalStartActionName:           n/a
VpnActionName:                  Silver-TransportMode
TunnelID:                       Y4
Type:                           Dynamic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Outbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        Deny
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  9.1.1.1
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    9.1.1.2
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
```

```
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   n/a
UpdateTime:                   n/a
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                1
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   Rule2Admin
FilterNameExtension:          1
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
SecurityClass:                0
Logging:                      Deny
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.1.1.1
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.2
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:49:48
UpdateTime:                   2012/02/14 11:07:20
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                1
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   Rule2Admin
FilterNameExtension:          2
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y4
Type:                         Dynamic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
```

```
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      Deny
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          n/a
SourceAddress:                9.1.1.2
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     n/a
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.1
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       n/a
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   n/a
UpdateTime:                   n/a
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                1
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   Rule2Admin
FilterNameExtension:          2
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      Deny
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.1.1.2
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   n/a
SourcePortRange:              n/a
SourcePortGranularity:        n/a
DestAddress:                  9.1.1.1
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     n/a
DestPortRange:                n/a
DestPortGranularity:          n/a
```

```
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:49:48
UpdateTime:                   2012/02/14 11:07:20
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                1
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*******************************************************************************

4 entries selected
```

NATT anchor and NATT dynamic filters

Using the **ipsec -f** command after the activation of two phase 2 Security Associations in the branch office with NAT model, the filter structure is shown in the following display:

```
CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 11:38:37 2012
Primary:  Filter          Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    Current           TotAvail: 139
Logging:  On              Predecap: Off               DVIPSec:  No
NatKeepAlive:  20         FIPS140:  No
Defensive Mode: Inactive

FilterName:                   Rule2C
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y2
Type:                         NATT Dynamic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.3.3.3
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.5.5.5
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 10.3.1.1
RmtUdpEncapPort:              4500
CreateTime:                   2012/02/14 10:19:52
UpdateTime:                   2012/02/14 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
```

```
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2C
FilterNameExtension:            1
GroupName:                      n/a
LocalStartActionName:           StartZoneC
VpnActionName:                  Gold-TunnelMode
TunnelID:                       Y3
Type:                           NATT Dynamic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Outbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.3.3.3
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     All
SourcePortRange:                n/a
SourcePortGranularity:          Rule
DestAddress:                    9.5.5.5
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       All
DestPortRange:                  n/a
DestPortGranularity:            Rule
OrigRmtConnPort:                n/a
RmtIDPayload:                   10.3.2.2
RmtUdpEncapPort:                4500
CreateTime:                     2012/02/14 10:19:52
UpdateTime:                     2012/02/14 10:19:52
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2C
FilterNameExtension:            1
GroupName:                      n/a
LocalStartActionName:           StartZoneC
VpnActionName:                  Gold-TunnelMode
TunnelID:                       Y0
Type:                           NATT Anchor
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Outbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
```

```
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.3.3.3
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     All
SourcePortRange:                n/a
SourcePortGranularity:          Rule
DestAddress:                    9.5.5.5
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       All
DestPortRange:                  n/a
DestPortGranularity:            Rule
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:19:52
UpdateTime:                     2012/02/14 10:19:52
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2C
FilterNameExtension:            1
GroupName:                      n/a
LocalStartActionName:           StartZoneC
VpnActionName:                  Gold-TunnelMode
TunnelID:                       Y0
Type:                           Dynamic Anchor
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Outbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.3.3.3
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     All
SourcePortRange:                n/a
SourcePortGranularity:          Rule
DestAddress:                    9.6.0.0
DestAddressPrefix:              16
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       All
DestPortRange:                  n/a
DestPortGranularity:            Rule
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:19:52
UpdateTime:                     2012/02/14 10:19:52
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
```

```
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2C
FilterNameExtension:            2
GroupName:                      n/a
LocalStartActionName:           StartZoneC
VpnActionName:                  Gold-TunnelMode
TunnelID:                       Y2
Type:                           NATT Dynamic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Inbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.5.5.5
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     All
SourcePortRange:                n/a
SourcePortGranularity:          Rule
DestAddress:                    9.3.3.3
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       All
DestPortRange:                  n/a
DestPortGranularity:            Rule
OrigRmtConnPort:                n/a
RmtIDPayload:                   10.3.1.1
RmtUdpEncapPort:                4500
CreateTime:                     2012/02/14 10:19:52
UpdateTime:                     2012/02/14 10:19:52
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  0
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2C
FilterNameExtension:            2
GroupName:                      n/a
LocalStartActionName:           StartZoneC
VpnActionName:                  Gold-TunnelMode
TunnelID:                       Y3
Type:                           NATT Dynamic
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Inbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
```

```
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.5.5.5
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 10.3.2.2
RmtUdpEncapPort:              4500
CreateTime:                   2012/02/14 10:19:52
UpdateTime:                   2012/02/14 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y0
Type:                         NATT Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.5.5.5
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:19:52
UpdateTime:                   2012/02/14 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
```

```
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y0
Type:                         Dynamic Anchor
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     All
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 n/a
ProtocolGranularity:          Rule
SourceAddress:                9.6.0.0
SourceAddressPrefix:          16
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   All
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     All
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              n/a
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2012/02/14 10:19:52
UpdateTime:                   2012/02/14 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************

8 entries selected
```

The inbound dynamic anchor filter protects TCP traffic from source address 9.6.0.0/16, source port any, to destination address 9.3.3.3, destination port 21. The inbound NATT anchor filter protects TCP traffic from source address 9.5.5.5, source port any, to destination address 9.3.3.3, destination port 21. The two inbound NATT dynamic filters also protect TCP traffic from source address 9.5.5.5, source port any, to destination address 9.3.3.3, destination port 21. However, the two NATT dynamic filters were negotiated for separate clients behind the security gateway. You can see that the first inbound NATT dynamic is for a host behind the security gateway using internal address 10.3.1.1 (value in the RmtIDpayload field).

The second inbound NATT dynamic is for a host behind the security gateway using internal address 10.3.2.2. The internal address of the data endpoint is what makes each NATT dynamic unique.

NAT resolution filters

Use the -h option on the **ipsec -f** command to display any NRFs associated with the displayed filters. After two clients behind the security gateway connect to host 9.3.3.3 using FTP, the NRFs might look like the following display. (The display is truncated to include only the NRFs.)

```
FilterName:                   Rule2C
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y2
Type:                         NRF
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     TCP(6)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 None
ProtocolGranularity:          Rule
SourceAddress:                9.3.3.3
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   21
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.5.5.5
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     34732
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              34732
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2010/02/16 10:19:52
UpdateTime:                   2010/02/16 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
*************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          1
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y3
Type:                         NRF
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Outbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     TCP(6)
ICMPType:                     n/a
```

```
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 None
ProtocolGranularity:          Rule
SourceAddress:                9.3.3.3
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   21
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.5.5.5
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     65535
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              34732
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2010/02/16 10:19:52
UpdateTime:                   2010/02/16 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y2
Type:                         NRF
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     TCP(6)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 None
ProtocolGranularity:          Rule
SourceAddress:                9.5.5.5
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   34732
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     21
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              34732
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2010/02/16 10:19:52
UpdateTime:                   2010/02/16 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
```

```
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
FilterName:                   Rule2C
FilterNameExtension:          2
GroupName:                    n/a
LocalStartActionName:         StartZoneC
VpnActionName:                Gold-TunnelMode
TunnelID:                     Y3
Type:                         NRF
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      All
LogLimit:                     n/a
Protocol:                     TCP(6)
ICMPType:                     n/a
ICMPTypeGranularity:          n/a
ICMPCode:                     n/a
ICMPCodeGranularity:          n/a
OSPFType:                     n/a
TCPQualifier:                 None
ProtocolGranularity:          Rule
SourceAddress:                9.5.5.5
SourceAddressPrefix:          n/a
SourceAddressRange:           n/a
SourceAddressGranularity:     Packet
SourcePort:                   65535
SourcePortRange:              n/a
SourcePortGranularity:        Rule
DestAddress:                  9.3.3.3
DestAddressPrefix:            n/a
DestAddressRange:             n/a
DestAddressGranularity:       Packet
DestPort:                     21
DestPortRange:                n/a
DestPortGranularity:          Rule
OrigRmtConnPort:              34732
RmtIDPayload:                 n/a
RmtUdpEncapPort:              n/a
CreateTime:                   2010/02/16 10:19:52
UpdateTime:                   2010/02/16 10:19:52
DiscardAction:                Silent
MIPv6Type:                    n/a
MIPv6TypeGranularity:         n/a
TypeRange:                    n/a
CodeRange:                    n/a
RemoteIdentityType:           n/a
RemoteIdentity:               n/a
FragmentsOnly:                No
FilterMatches:                0
LifetimeExpires:              n/a
AssociatedStackCount:         n/a
************************************************************************
```

There are two NRF inbound/outbound entry pairs associated with the NATT anchor. In this example, two clients behind the security gateway have an FTP connection with host 9.3.3.3. The first outbound NRF entry is for:

```
source address 9.3.3.3, source port 21
destination address 9.5.5.5, destination port 34732
protocol TCP
```

The destination port is shown in the DestPort field. This value can be a translated value. The OrigRmtConnPort field indicates the original remote connection port, prior to remote port translation

by Communications Server. In this example, the first outbound NRF shows that DestPort and OrigRmtConnPort are both 34732. For more information, see “Remote port translation” on page 934.

The second outbound NRF entry is for:

```
source address 9.3.3.3, source port 21
destination address 9.5.5.5, destination port 65535
protocol TCP
```

The original remote connection port (OrigRmtConnPort) is 34732. Because the values in DestPort and OrigRmtConnPort do not match, you can tell that the value was translated by Communications Server's remote port translation function. For more information, see “Remote port translation” on page 934.

The TunnelID field provides information on which phase 2 Security Association the traffic will be sent over. In this example, the phase two Security Associations are identified by the labels Y2 and Y3 respectively.

Displaying remote port translation with the ipsec command

As seen in “NAT resolution filters” on page 1062, the remote data endpoint is represented by the security gateway's public IP address (9.5.5.5), not the client's IP address (10.3.1.1 or 10.3.2.2). Using the **ipsec -o display** command after the activation of two FTP connections in the branch office with NAT model, the port mappings are shown in the following display:

```
CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:43:55 2010
Primary:  NATT Port Trans Function: Display           Format:   Detail
Source:   Stack           Scope:    Current           TotAvail: 2

RmtIpAddress:         9.5.5.5
Protocol:             TCP(6)
TransRmtConnPort:     34732
OrigRmtConnPort:      34732
RmtInnerIpAddress:    10.3.1.1
*********************************************************************
RmtIpAddress:         9.5.5.5
Protocol:             TCP(6)
TransRmtConnPort:     65535
OrigRmtConnPort:      34732
RmtInnerIpAddress:    10.3.2.2
*********************************************************************

2 entries selected
```

In both entries, you can see that the remote IP address (RmtIpAddress) value is 9.5.5.5, the IP address of the branch office gateway, the protocol is TCP (6), and the original remote connection port (OrigRmtConnPort) is 34732. The first entry shows that the translated remote connection port (TransRmtConnPort) is also 34732. The remote inner IP address contains the private address of the client behind the security gateway that initiated the connection, 10.3.1.1. The second entry shows that the remote connection port was translated to a value of 65535 (TransRmtConnPort), and that the client initiating the connection is using private IP address 10.3.2.2.

Table 53 on page 1065 details several places where one or both of the remote port values are displayed or used for a selection.

Table 53. Original and translated port values

Function	How remote port values are used	Which remote port, original or translated?
Netstat displays of connection data, such as Netstat ALL/-A, Netstat ALLConn/-a, and Netstat COnn/-c	The Netstat command has many options to display connection information, including the remote port value. In some cases, the Netstat command takes a remote port value as a selector.	Translated remote port

Table 53. Original and translated port values (continued)

Function	How remote port values are used	Which remote port, original or translated?
Netstat display of the VIPA connection routing table (netstat VCRT/-V)	This command displays the remote port value in the sport or Source field, depending on the flavor of the report generated, and allows you to select based on port.	Translated remote port
Packet trace	Packet trace displays packet data as it was received or sent. If the packet is authenticated but not encrypted, the port is visible in the packet trace data.	Original remote port
IPSecurity syslog messages: • `EZD0814I packet permitted` • `EZD0815I packet denied by policy` • `EZD0821I packet denied, no tunnel` • `EZD0822I packet denied, tunnel inactive` • `EZD0832I packet denied by NAT traversal processing` • `EZD0833I packet denied, tunnel mismatch` • `EZD0836I packet permitted` Defensive filtering syslog messages: • `EZD1721I packet denied by defensive filter` • `EZD1722I packet would have been denied by defensive filter`	For an inbound packet, the sport field in these messages contains a remote port value. For an outbound packet, the dport field in these messages contains a remote port value. These messages also have an origport field.	The sport and dport fields contain the translated remote port, and the origport field contains the original remote port.
Dynamic anchor, displayed with the **`ipsec -f`** command and the **`ipsec -t`** command	The dynamic anchor that is configured in the Policy Agent configuration file can specify the remote port as a single port, a range of ports, or all ports. This specification of the remote port controls the range of ports that the original port can be translated to. Both the original port and the translated port for a connection will fit the range of ports coded.	The configured remote port value is displayed. A packet's original port is used to match on this rule. Both the original port and translated port are included in the remote port value displayed.
NATT anchor, displayed with the **`ipsec -f`** command and the **`ipsec -t`** command	The NATT anchor, which is created as a result of the Security Association negotiation, contains a specific remote port or all ports.	Original remote port if specific port displayed

Table 53. Original and translated port values (continued)		
Function	**How remote port values are used**	**Which remote port, original or translated?**
NATT dynamic, displayed with the **ipsec -f** command and the **ipsec -t** command	The NATT dynamic, which is created as a result of the Security Association negotiation, contains a specific remote port or all ports.	Original remote port if specific port displayed
NAT resolution filter (NRF), displayed with the **ipsec -f** command with the -h option	The NAT resolution filter is a connection level filter, and contains both the translated remote port and the original remote port.	Translated remote port and original remote port. A packet's translated port is used to match on this rule.
ipsec -t command	The **ipsec** traffic test command allows a remote port value to be specified as a filter selection criteria.	Original remote port

Displaying Security Associations with the ipsec command

Use the **ipsec** command to verify Security Associations.

Displaying IKE tunnel information with the ipsec command

Use the **ipsec -k display** command to display IKE tunnel information.

```
ipsec -k display

CS V2R5 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:25 2020
Primary:  IKE tunnel      Function: Display            Format:   Detail
Source:   IKED            Scope:    Current            TotAvail: n/a

TunnelID:                     K3
Generation:                   1
IKEVersion:                   1.0
KeyExchangeRuleName:          ZoneC_KeyExRule1
KeyExchangeActionName:        Gold-PSK
LocalEndPoint:                9.3.3.3
LocalIDType:                  IPV4
LocalID:                      9.3.3.3
RemoteEndPoint:               10.3.1.1
RemoteIDType:                 USERFQDN
RemoteID:                     gateway.poughkeepsie.ibm.com
ExchangeMode:                 Main
State:                        DONE
AuthenticationAlgorithm:      HMAC-MD5
EncryptionAlgorithm:          3DES-CBC
 KeyLength:                   n/a
PseudoRandomFunction:         HMAC-MD5
DiffieHellmanGroup:           2
LocalAuthenticationMethod:    PresharedKey
RemoteAuthenticationMethod:   PresharedKey
InitiatorCookie:              0xE70D94ADB3D75947
ResponderCookie:              0xCED4B800A0BE81BC
Lifesize:                     0K
CurrentByteCount:             296b
Lifetime:                     480m
LifetimeRefresh:              2020/02/16 19:15:22
LifetimeExpires:              2020/02/16 19:23:19
ReauthInterval:               480m
ReauthTime:                   2020/02/16 19:15:22
Role:                         Responder
AssociatedDynamicTunnels:     2
NATTSupportLevel:             RFC
NATInFrntLclScEndPnt:         No
NATInFrntRmtScEndPnt:         Yes
zOSCanInitiateP1SA:           Yes
AllowNat:                     Yes
RmtNAPTDetected:              No
RmtUdpEncapPort:              4500
```

```
LocalCertExpires:             n/a
LocalSerialNumber:            n/a
LocalIssuerDNLength:          0
LocalIssuerDN:                n/a
LocalSubjectDNLength:         0
LocalSubjectDN:               n/a
RemoteCertExpires:            n/a
RemoteSerialNumber:           n/a
RemoteIssuerDNLength:         0
RemoteIssuerDN:               n/a
RemoteSubjectDNLength:        0
RemoteSubjectDN:              n/a
*********************************************************************

1 entries selected
```

The setting of the AllowNat field indicates whether or not NAT traversal support was advertised to the IKE peer. If AllowNat is Yes, the negotiation might or might not have detected a NAT. If the NATInFrntLclScEndPnt field is Yes, a NAT device was detected in front of the local security endpoint. If the NATInFrntRmtScEndPt field is Yes, a NAT device was detected in front of the remote security endpoint.

Displaying IPSec tunnel information with the ipsec command

Use the **ipsec -y display** command to display IPSec tunnel information.

```
ipsec -y display -a Y39

TunnelID:                     Y39
Generation:                   1
IKEVersion:                   1.0
ParentIKETunnelID:            K11
VpnActionName:                TransportMode
LocalDynVpnRule:              n/a
State:                        Active
HowToEncap:                   Transport
LocalEndPoint:                9.2.2.2
RemoteEndPoint:               9.4.4.4
LocalAddressBase:             9.2.2.2
LocalAddressPrefix:           n/a
LocalAddressRange:            n/a
RemoteAddressBase:            9.4.4.4
RemoteAddressPrefix:          n/a
RemoteAddressRange:           n/a
HowToAuth:                    ESP
 AuthAlgorithm:               HMAC-MD5
 AuthInboundSpi:              2418545801 (0x90281489)
 AuthOutboundSpi:             4027602341 (0xF01055A5)
HowToEncrypt:                 DES-CBC
 KeyLength:                   n/a
 EncryptInboundSpi:           2418545801 (0x90281489)
 EncryptOutboundSpi:          4027602341 (0xF01055A5)
Protocol:                     ALL(0)
LocalPort:                    n/a
LocalPortRange:               n/a
RemotePort:                   n/a
RemotePortRange:              n/a
Type:                         n/a
TypeRange:                    n/a
Code:                         n/a
CodeRange:                    n/a
OutboundPackets:              1
OutboundBytes:                264
InboundPackets:               1
InboundBytes:                 264
Lifesize:                     0K
LifesizeRefresh:              0K
CurrentByteCount:             0b
LifetimeRefresh:              2010/02/16 15:14:52
LifetimeExpires:              2010/02/16 15:23:19
CurrentTime:                  2010/02/16 11:53:31
VPNLifeExpires:               2010/02/17 11:23:19
NAT Traversal Topology:
  UdpEncapMode:               Yes
  LclNATDetected:             No
  RmtNATDetected:             Yes
  RmtNAPTDetected:            No
  RmtIsGw:                    No
  RmtIsZOS:                   Yes
```

```
  zOSCanInitP2SA:             Yes
  RmtUdpEncapPort:            4500
  SrcNATOARcvd:               10.2.2.2
  DstNATOARcvd:               9.2.2.2
PassthroughDF:                n/a
PassthroughDSCP:              n/a
*********************************************************************

1 entries selected
```

The NAT Traversal Topology fields show additional information when a NAT was detected in the path between the IKE peers. The setting of the UdpEncapMode field indicates whether a UDP-encapsulated mode Security Association has or has not been negotiated. If NAT Traversal is supported by both IKE peers and one or more NATs are detected, UdpEncapMode is set to Yes. The RmtNATDetected field is Yes if a NAT is detected in front of the remote peer. The RmtIsGW field is Yes if the remote peer is acting as a security gateway.

Tip: Use the -b option of the **ipsec -y display** command to show the ports and protocols of the dynamic filter that are associated with the phase 2 Security Association. The following excerpt from the **ipsec -y display** using the -b option indicates a Telnet connection from a remote host:

```
AssociatedFiltProtocol:     TCP(6)
AssociatedFiltSrcPort:      23
AssociatedFiltDestPort:     0
```

Displaying filter rules with the pasearch command

The configured IP filter rules and associated actions can also be viewed from the perspective of the Policy Agent. The **pasearch** command provides a way to view all Policy Agent configuration, of which IP security is a subset. In contrast to the information that is provided by the **ipsec** command, the detailed information that is provided by the **pasearch** command does not reflect the stack's active use of the IP security policy, but offers a relatively static view of configured IP security values that were generated from the IP security configuration file. For more information on displaying policy based networking information, see z/OS Communications Server: IP System Administrator's Commands.

Verifying filter action

To quickly determine which filter rule applies to a specific type of traffic, use the **ipsec** traffic test command (**ipsec -t**). This command returns all of the rules in the current filter table that match the given traffic type.

For example, to test which filter rule matches an incoming FTP connection request from remote IP address 9.1.1.2 to local IP address 9.1.1.1, issue the following command. The input values represent the remote address, local address, protocol, remote port, local port, direction, and security class of the packet.:

```
ipsec -t  9.1.1.2 9.1.1.1 tcp 0 21 in 0

CS V2R1 ipsec  Stack Name: TCPCS  Tue Feb 14 11:59:45 2012
Primary:  IP Traffic Test Function: Display           Format:   Detail
Source:   Stack Policy    Scope:    n/a               TotAvail: 5
TestData: 9.1.1.2  9.1.1.1  tcp 0 21 in 0
Defensive Mode: Inactive

FilterName:                   Rule2Admin
FilterNameExtension:          2
GroupName:                    Admin
LocalStartActionName:         n/a
VpnActionName:                Silver-TransportMode
TunnelID:                     Y4
Type:                         Dynamic
DefensiveType:                n/a
State:                        Active
Action:                       Permit
Scope:                        Local
Direction:                    Inbound
OnDemand:                     No
SecurityClass:                0
Logging:                      Deny
```

```
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  9.1.1.2
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    9.1.1.1
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     n/a
UpdateTime:                     n/a
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  1
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
*************************************************************************
FilterName:                     Rule2Admin
FilterNameExtension:            2
GroupName:                      Admin
LocalStartActionName:           n/a
VpnActionName:                  Silver-TransportMode
TunnelID:                       Y0
Type:                           Dynamic Anchor
DefensiveType:                  n/a
State:                          Active
Action:                         Permit
Scope:                          Local
Direction:                      Inbound
OnDemand:                       No
SecurityClass:                  0
Logging:                        Deny
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            Rule
SourceAddress:                  9.1.1.2
SourceAddressPrefix:            n/a
SourceAddressRange:             n/a
SourceAddressGranularity:       Packet
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    9.1.1.1
DestAddressPrefix:              n/a
DestAddressRange:               n/a
DestAddressGranularity:         Packet
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:49:48
```

```
UpdateTime:                    2012/02/14 11:07:20
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 1
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
************************************************************************
FilterName:                    Rule1All-IPv4-Permit
FilterNameExtension:           6
GroupName:                     ZoneAll
LocalStartActionName:          n/a
VpnActionName:                 n/a
TunnelID:                      0x00
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Permit
Scope:                         Local
Direction:                     Inbound
OnDemand:                      n/a
SecurityClass:                 0
Logging:                       None
LogLimit:                      n/a
Protocol:                      TCP(6)
ICMPType:                      n/a
ICMPTypeGranularity:           n/a
ICMPCode:                      n/a
ICMPCodeGranularity:           n/a
OSPFType:                      n/a
TCPQualifier:                  Connect Outbound
ProtocolGranularity:           n/a
SourceAddress:                 0.0.0.0
SourceAddressPrefix:           0
SourceAddressRange:            n/a
SourceAddressGranularity:      n/a
SourcePort:                    53
SourcePortRange:               n/a
SourcePortGranularity:         n/a
DestAddress:                   0.0.0.0
DestAddressPrefix:             0
DestAddressRange:              n/a
DestAddressGranularity:        n/a
DestPort:                      All
DestPortRange:                 n/a
DestPortGranularity:           n/a
OrigRmtConnPort:               n/a
RmtIDPayload:                  n/a
RmtUdpEncapPort:               n/a
CreateTime:                    2012/02/14 10:49:48
UpdateTime:                    2012/02/14 10:49:48
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 0
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
************************************************************************
FilterName:                    Rule2All-IPv4-Deny
FilterNameExtension:           2
GroupName:                     ZoneAll
LocalStartActionName:          n/a
VpnActionName:                 n/a
TunnelID:                      0x00
Type:                          Generic
DefensiveType:                 n/a
State:                         Active
Action:                        Deny
Scope:                         Both
Direction:                     Inbound
OnDemand:                      n/a
SecurityClass:                 0
```

```
Logging:                        All
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
CreateTime:                     2012/02/14 10:49:48
UpdateTime:                     2012/02/14 10:49:48
DiscardAction:                  Silent
MIPv6Type:                      n/a
MIPv6TypeGranularity:           n/a
TypeRange:                      n/a
CodeRange:                      n/a
RemoteIdentityType:             n/a
RemoteIdentity:                 n/a
FragmentsOnly:                  No
FilterMatches:                  40
LifetimeExpires:                n/a
AssociatedStackCount:           n/a
************************************************************************
FilterName:                     DenyAllRule_Generated__________Inbnd
FilterNameExtension:            n/a
GroupName:                      n/a
LocalStartActionName:           n/a
VpnActionName:                  n/a
TunnelID:                       0x00
Type:                           Generic
DefensiveType:                  n/a
State:                          Active
Action:                         Deny
Scope:                          Both
Direction:                      Inbound
OnDemand:                       n/a
SecurityClass:                  0
Logging:                        None
LogLimit:                       n/a
Protocol:                       All
ICMPType:                       n/a
ICMPTypeGranularity:            n/a
ICMPCode:                       n/a
ICMPCodeGranularity:            n/a
OSPFType:                       n/a
TCPQualifier:                   n/a
ProtocolGranularity:            n/a
SourceAddress:                  0.0.0.0
SourceAddressPrefix:            0
SourceAddressRange:             n/a
SourceAddressGranularity:       n/a
SourcePort:                     n/a
SourcePortRange:                n/a
SourcePortGranularity:          n/a
DestAddress:                    0.0.0.0
DestAddressPrefix:              0
DestAddressRange:               n/a
DestAddressGranularity:         n/a
DestPort:                       n/a
DestPortRange:                  n/a
DestPortGranularity:            n/a
OrigRmtConnPort:                n/a
RmtIDPayload:                   n/a
RmtUdpEncapPort:                n/a
```

```
CreateTime:                    2012/02/14 10:36:09
UpdateTime:                    2012/02/14 10:49:48
DiscardAction:                 Silent
MIPv6Type:                     n/a
MIPv6TypeGranularity:          n/a
TypeRange:                     n/a
CodeRange:                     n/a
RemoteIdentityType:            n/a
RemoteIdentity:                n/a
FragmentsOnly:                 No
FilterMatches:                 0
LifetimeExpires:               n/a
AssociatedStackCount:          n/a
*******************************************************************************

5 entries selected
```

An incoming FTP connection request matches all of the rules shown in the example. The first rule that is returned does not always match a specific packet, depending on how much detail you provide as the input to the **ipsec -t** command. However, the Rule2Admin rule is the best match in this case, so the search for a matching filter ends there. The matching rule in this case is an ipsec rule, as indicated by the designation `Dynamic Anchor`. Therefore, IPSec processing is applied to this packet.

Tip: When using the **ipsec -t** command, provide as much detailed input as possible. The more detailed the input to the command, the more narrow the results of the search will be.

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Security Associations

This topic includes the following subtopics:

- Activating a Security Association
- Verifying the activation of a Security Association
- Verifying the use of an active Security Association
- Refreshing Security Associations
- Deactivating Security Associations

Activating a Security Association

Negotiations can be initiated in one of four ways:

- Remote activation

  When a remote IKE peer initiates a negotiation with the local IKE daemon, no action is required. If the IP security policy has been configured correctly and is consistent with the policy of the remote IKE peer, a Security Association is established. No operator message is issued when a remote activation has occurred, but the syslog does contain a record of all IKE activity. The **ipsec -y display** command can also be used to view all of the active Security Associations.

- On-demand activation

  An on-demand Security Association is activated when some outbound traffic matches an ipsec rule that allows on-demand activation. The `ondemand` field of the filter display indicates whether or not on-demand activation is allowed for that rule.

- Automatic activation

  The local IKE daemon initiates a negotiation for an autoactivated Security Association when it connects to the TCP/IP stack. IKE also initiates a negotiation for an autoactivated Security Association when the **ipsec -f reload** command is issued, changing the active filter rule set from default IP filter rules to Policy Agent filter rules. No operator message is issued when an autoactivation has occurred, but the syslog does contain a record of all IKE activity. The **ipsec -y display** command can also be used to view all of the active Security Associations.

- Command-line activation

  The **ipsec** command can be used as follows to activate a Security Association that has been defined by a LocalDynVpnRule statement:

```
ipsec -y activate -l ZoneC_VPN-EE1

CS V1R12 ipsec  Stack Name: TCPCS  Wed Feb  3 16:02:05 2010
Primary:  Dynamic tunnel  Function: Activate

Selection Data                                  Status
ZoneC_VPN-EE1                                   Activating
```

  The output of the command indicates the status of the activation.

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Verifying the activation of a Security Association

After a Security Association has been activated, it can be displayed with the **ipsec -y display** command. To view all active Security Associations, issue the following command:

```
ipsec -y display
```

You can use the **ipsec** command to view all of the active phase 1 Security Associations, or limit the report to a single phase 1 Security Association by using the -a option.

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Verifying the use of an active Security Association

If filter logging is enabled for the selected filter rule, the log indicates whether a packet has been permitted with IPSec processing applied. Among the information available for a typical filter log entry are the rule name, the action, and the tunnel ID:

```
Feb 13 18:09:11 MVS175/TRMD     TRMD1    TRMD.TCPCS2[28]: EZD0814I Packet
permitted: 02/13/2010 18:09:05.96 filter rule= Rule2Admin ext= 1 sipaddr=
9.1.1.2 dipaddr= 9.1.1.1 proto= tcp(6) sport= 3755 dport= 21 -=
Interface= 9.1.1.1 (I) secclass= 255 dest= local len= 52 vpnaction=
Silver-TransportMode tunnelID= Y58 ifcname= MPC4142L fragment= N
```

The **ipsec -y display** command also outputs a field with the number of bytes of traffic that have been protected by a particular Security Association.

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Refreshing Security Associations

When a Security Association is refreshed, the encryption keys change. Refreshing a Security Association periodically prevents the keys from being compromised by an outside party. Phase 1 and phase 2 Security Associations are refreshed automatically, based on the lifetime or life size that was configured for IKEv2 or negotiated between the two IKE peers for IKEv1. When a lifetime expiration causes an IKEv2 phase 1 Security Association to refresh, the encryption key changes but the peer is not reauthenticated. Changing the key without reauthenticating the peer reduces CPU cost.

Tip: For phase 1, these parameters are specified in the KeyExchangeOffer statement. For phase 2, these parameters are specified in the IpDataOffer statement.

You can also refresh Security Associations from the z/OS UNIX command line, but this should be necessary only in exceptional conditions because the IKE daemon is normally responsible for refreshing the keys at configured intervals. Exceptional conditions might include the compromise of a key or the failure to receive an informational IKE message from a remote host. For both IKEv1 and IKEv2 Phase

1 Security Associations, refreshes from the z/OS UNIX command line include both reauthentication and re-keying.

Phase 1

Each phase 1 Security Association is identified by a tunnel ID, a number with a prefix of *K*. To manually refresh a phase 1 Security Association, issue the **ipsec -k display** command to find the tunnel ID. Then issue the **ipsec -k refresh** command for that ID as follows:

```
ipsec -k refresh -a K1

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  IKE tunnel      Function: Refresh

Tunnel ID    Status
K1           Refreshing
```

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Phase 2

Each phase 2 Security Association is identified by a tunnel ID, a number with a prefix of *Y*. To manually refresh a phase 2 Security Association, issue the **ipsec -y display** command to find the tunnel ID. Then issue the **ipsec -y refresh** command for that ID as follows:

```
ipsec -y refresh -a Y2

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  Dynamic tunnel  Function: Refresh

Tunnel ID    LocalDynVpnRuleName                              Status
Y2           ZoneC_VPN-EE1                                    Refreshing
```

The phase 2 Security Association can also be identified by the local dynamic VPN rule with which it is associated, if one exists, as follows:

```
ipsec -y refresh -l ZoneC_VPN-EE1

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  Dynamic tunnel  Function: Refresh

Tunnel ID    LocalDynVpnRuleName                              Status
Y2           ZoneC_VPN-EE1                                    Refreshing
```

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Deactivating Security Associations

When a Security Association is deleted, all of the information that is stored in the Security Association is deleted from the TCP/IP stack and from the IKED, along with the dynamic filters that were created when the Security Association was created. After deletion, the Security Association is no longer available for use. Traffic that was protected by the old Security Association is denied until a new Security Association is subsequently activated.

When a parent phase 1 Security Association is deactivated, all of the associated phase 2 Security Associations are deleted as well. Be careful when deleting phase 1 Security Associations, because all traffic that uses the Security Association and its associated phase 2 Security Associations are dropped until new Security Associations can be negotiated.

- To delete a phase 1 Security Association and all of the phase 2 Security Associations it is protecting, issue the following command:

  ```
  ipsec -k deactivate -a K1
  ```

```
CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  IKE tunnel      Function: Deactivate

Tunnel ID    Status
K1           Deactivating
```

- To delete all phase 1 Security Associations and all phase 2 Security Associations, use the **-a all** option as follows:

```
ipsec -k deactivate -a all

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  IKE tunnel      Function: Deactivate

All IKE tunnels Deactivating
```

- To delete a phase 2 Security Association, issue the following command:

```
ipsec -y deactivate -a Y2

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  Dynamic tunnel  Function: Deactivate

Tunnel ID    LocalDynVpnRuleName                              Status
Y2           n/a                                              Deactivating
```

  The n/a in the LocalDynVpnRuleName field indicates that no local dynamic VPN rule name is associated with this Security Association. The Security Association was either remotely activated or was activated on-demand.

- To delete all phase 2 Security Associations, use the **-a all** option as follows:

```
ipsec -y deactivate -a all

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 11:48:04 2010
Primary:  Dynamic tunnel  Function: Deactivate

All dynamic tunnels Deactivating
```

For detailed information about the use of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Modifying active IP security policy

This topic describes the effects of changes to security-related files and of issuing the **ipsec -f default** command.

IP security policy files

The IP security configuration files for a TCP/IP stack that has the IPSECURITY parameter defined can be modified while the stack is active in the following ways:

- Policy Agent configuration, including IP security, is updated at each configured refresh interval that is specified on the TcpImage statement for local policies or on the DynamicConfigPolicyLoad statement for remote policies. The default is 30 minutes.
- If Policy Agent was started with the `-i` option and the configuration files are stored in a z/OS UNIX file, changes to any of the configuration files are detected and updated dynamically without user intervention. The `-i` option has no effect when the policies are stored in remote configuration files on the policy server.
- Policy Agent configuration, including IP security, can be updated at any time by issuing the MODIFY PAGENT,REFRESH command or the MODIFY PAGENT,UPDATE command from the console of the policy client.

Policy Agent image configuration files

For an active TCP/IP stack with IPSECURITY defined, the following conditions apply:

- Changing the file name that is identified by an existing IpSecConfig or DynamicConfigPolicyLoad statement causes Policy Agent to update and install the IP security policy as defined in the new file. If the new IP security policy file contains errors, the policy is not updated and the existing IP security policy remains in effect.
- For local IP security policies, removing an existing IpSecConfig statement from a Policy Agent image configuration file activates the default IP filter policy.
- For remote IP security policies, removing the PolicyServer statement (or the PolicyType IPSec parameter on that statement) from a Policy Agent image configuration file activates the default IP filter policy, assuming that no local IP security policy is defined using the IpSecConfig statement.

Policy Agent main configuration file

For an active TCP/IP stack with IPSECURITY defined, removing an existing TcpImage statement from the Policy Agent configuration file has the following effects on IP security policy:

- Existing IP filters remain.
- Existing Security Associations remain.
- Traffic continues to flow in the same way it did before the TcpImage statement was removed, including IPSec-protected traffic.
- New Security Associations cannot be activated.
- Existing Security Associations cannot be refreshed and are deleted when the refresh period expires.

If the intent of removing the TcpImage statement is to remove IP filters from the stack, an alternative is to modify the IP security policy to install a filter rule that permits all inbound and outbound traffic. Also, before restarting the stack, the IPSECURITY parameter should be removed from the IPCONFIG statement of the relevant stack.

Active Security Associations and the ipsec -f default command

Any active Security Associations that were negotiated for IPSec-protected traffic are not deleted when the **ipsec -f default** command is issued. However, they are deleted if, while the default policy is in effect, any associated IP filter rules from the IP filter policy are deleted or modified in such a way that the filter rule no longer encompasses the scope of the Security Association. In that case, the Security Association will be deleted when the IP security policy is reloaded.

For example, Security Associations are not deleted by the following sequence of actions:

1. The **ipsec -f default** command is issued.

   Security Associations remain active in the stack and in IKE, though unavailable for use.
2. No modification is made to the IP filter policy in the IP security configuration files.

   Security Associations remain active in the stack and in IKE, though unavailable for use.
3. The **ipsec -f reload** command is issued.

   Security Associations remain active in the stack and in IKE, and are available for use.

Security Associations are deleted by the following sequence of actions:

1. The **ipsec -f default** command is issued.

   Security Associations remain active in the stack and in IKE.
2. The IpFilterRule statement that is associated with an active Security Association is deleted.
3. The IP security policy is updated by issuing the MODIFY PAGENT,REFRESH command from the console.

   Existing Security Associations are deleted.
4. The **ipsec -f reload** command is issued.

   Security Associations have been deleted.

In either case, Security Associations are never available for use when the default IP filter policy is in effect.

Displaying NSS client information

Use the NssStackConfig statement to configure a stack as an NSS client. Use the `-w` primary option on the **ipsec** command to determine which active stacks are configured as NSS clients, as well as their current status.

```
ipsec -w display

CS V1R12 ipsec  NSS Client Name: n/a  Tue Feb 16 12:14:24 2010
Primary:  Stack NSS       Function: Display            Format:   Detail
Source:   IKED            Scope:    n/a                TotAvail: 3
SystemName: MVS175

StackName:                      TCPCS
ClientName:                     n/a
ClientAPIVersion:               n/a
ServerAPIVersion:               n/a
NSServicesSupported:            No
RemoteManagementSelected:       No
RemoteManagementEnabled:        n/a
CertificateServicesSelected:    No
CertificateServicesEnabled:     n/a
NSClientIPAddress:              n/a
NSClientPort:                   n/a
NSServerIPAddress:              n/a
NSServerPort:                   n/a
NSServerSystemName:             n/a
UserID:                         n/a
ConnectionState:                n/a
TimeConnectedToNSServer:        n/a
TimeOfLastMessageToNSServer:    n/a
*********************************************************************
StackName:                      TCPCS4
ClientName:                     Client4
ClientAPIVersion:               4
ServerAPIVersion:               4
NSServicesSupported:            Yes
RemoteManagementSelected:       Yes
RemoteManagementEnabled:        Yes
CertificateServicesSelected:    Yes
CertificateServicesEnabled:     Yes
NSClientIPAddress:              10.81.4.4
NSClientPort:                   50008
NSServerIPAddress:              10.81.5.5
NSServerPort:                   4159
NSServerSystemName:             MVS175
UserID:                         USER1
ConnectionState:                connected
TimeConnectedToNSServer:        2010/02/16 12:12:42
TimeOfLastMessageToNSServer:    2010/02/16 12:12:45
*********************************************************************
StackName:                      TCPCS5
ClientName:                     V1RCIPSECREG_TCPCS5_SGWR
ClientAPIVersion:               4
ServerAPIVersion:               4
NSServicesSupported:            Yes
RemoteManagementSelected:       Yes
RemoteManagementEnabled:        Yes
CertificateServicesSelected:    Yes
CertificateServicesEnabled:     Yes
NSClientIPAddress:              10.81.5.5
NSClientPort:                   50000
NSServerIPAddress:              10.81.5.5
NSServerPort:                   4159
NSServerSystemName:             MVS175
UserID:                         USER1
ConnectionState:                connected
TimeConnectedToNSServer:        2010/02/16 12:11:59
TimeOfLastMessageToNSServer:    2010/02/16 12:11:59
*********************************************************************

3 entries selected
```

For complete details about the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Sysplex-Wide Security Associations and IP security

z/OS IP security supports Sysplex-Wide Security Associations (SWSA). In a sysplex environment, SWSA distributes IPSec Security Associations (SAs) to the target stacks of distributed DVIPAs. To enable this support for IPv4 DVIPAs, you must code IPCONFIG IPSECURITY in the TCP/IP profile and DVIPSEC in the IPSEC block. To enable this support for IPv6 DVIPAs, you must code IPCONFIG6 IPSECURITY, in addition to IPCONFIG IPSECURITY and DVIPSEC.

You should add DVLOCALFLTR, the DVIPSEC subparameter, to the IPSEC statement block of the TCP/IP profile when intra-sysplex traffic should be protected by security associations. DVLOCALFLTR enables IP filtering and IPSec protection of TCP traffic between a client and an IPv4 dynamic VIPA that are defined on the same TCP/IP stack, when the traffic is forwarded to another TCP/IP stack. When the DVLOCALFLTR parameter is configured, ensure that the IPSec policy accounts for all local TCP traffic with an IPv4 dynamic VIPA endpoint. Traffic that does not match a configured IP filter rule is denied.

Restrictions:

- The IKEv1 protocol cannot be used to negotiate a tunnel between a client and an IPv4 dynamic VIPA that are defined on the same TCP/IP stack. You must use the IKEv2 protocol to negotiate a tunnel to protect the traffic. Use HowToInitiate IKEv2 on the KeyExchangePolicy statement or a specific KeyExchangeAction statement to indicate that the IKEv2 protocol must be used when this system starts key negotiations.
- A tunnel cannot be negotiated if the source and destination IP addresses for client connections are the same dynamic VIPA. To help avoid this scenario, do not code a dynamic VIPA that can be used as a destination IP address on the TCPCONFIG TCPSTACKSOURCEVIPA or SRCIP statement in the TCP/IP profile.

In a sysplex environment, the IKE daemon can detect movement of a DVIPA to or from an IP security stack. To reestablish the Security Associations of a DVIPA, the DVIPSEC parameter must be specified in the TCP/IP profile of the stack from which the DVIPA is being moved and in the TCP/IP profile of the stack to which the DVIPA is being moved. For details about this movement, see “Sysplex-Wide Security Associations” on page 396. For information about the IPCONFIG, IPCONFIG6, and IPSEC statements that need to be added to the TCP/IP profile to configure this support, see z/OS Communications Server: IP Configuration Reference.

When a DVIPA is moved from one IP security stack in a sysplex to another IP security stack, and both stacks have the DVIPSEC option specified, an attempt is made to automatically reestablish Security Associations on the backup stack. The IKE daemon on the system that is assuming control of the DVIPA attempts to renegotiate new Security Associations to replace the ones that were on the system that previously owned the DVIPA. If these attempts fail due to configuration errors or connectivity errors, manual intervention might be required. Phase 1 Security Association or phase 2 Security Association negotiations that were in progress at the time of the DVIPA movement are lost. However, if these negotiations were for a refresh, a new negotiation is started in the process of assuming control of the DVIPA.

When a DVIPA is moved from one IP security stack in a sysplex to another IP security stack, and one or both stacks do not have the DVIPSEC option specified, the Security Associations that are associated with that DVIPA must be reestablished by issuing the **ipsec** command, on-demand activation, or by a peer initiation.

Guidelines:

- If a DVIPA is manually deleted and that address has no backup, the IKE daemon might not be able to terminate the tunnels in which the DVIPA is a security endpoint. To avoid this problem, use the **ipsec** command to deactivate the DVIPA's IKE tunnels before manually deleting the DVIPA.
- This support does not address the dynamic relocation of static filter rules and VPN policy definitions to the target system in the sysplex. It is up to you to ensure that the necessary filter rules and IP security policy definitions exist on all participating systems in the sysplex. If the necessary filter rules and IP

security policy definitions do not exist, the IKE daemon might not be able to reestablish all Security Associations. For a description of the SWSA function, see "Sysplex-Wide Security Associations" on page 396.

Rule: Because the renegotiation of a Security Association after a DVIPA move requires the sysplex stack to initiate an IKE negotiation, the sysplex stack must be allowed to initiate. You must code the Initiation attribute on the IpDynVpnAction statement as `localonly` or `either`.

NAT traversal and Sysplex-Wide Security Associations

The following topics describe Sysplex-Wide Security Associations (SWSA) in NAT traversal (NATT) configurations.

AES-GCM

When the AES-GCM combined-mode encryption and authentication algorithm is negotiated for an SA, decapsulation of all packets over that SA occurs on the distributing stack, and the decapsulated packets are forwarded to the target stack.

DVIPA recovery support

In NATT configurations where IKE can act only as the responder, sysplex distribution is possible but the recovery of the Security Associations when the DVIPA moves is not supported. There are two NATT configurations in which IKE can act only as the responder:

- When the IKE peer is a security gateway and a NAT is being traversed
- When the IKE peer is behind a NAPT

For more information about NATT configurations, as well as interoperability considerations, see "Configuration scenarios supported for NAT traversal" on page 1036.

When only one client behind a NAPT has negotiated a Security Association, it is not always possible for the server to detect whether one-to-one address translation or many-to-one address port translation (NAPT) is being done. When multiple clients have active Security Associations, the server can detect that port translation is being done. If z/OS cannot determine that a Security Association is negotiated with a remote peer behind a NAPT, the Security Association is treated as if it is being negotiated with a remote peer using one-to-one address translation.

When IKE is limited to only a responder role, the Security Association must be reestablished by peer initiation. The interoperability considerations for establishing an initial phase 2 Security Association are relevant to the renegotiation of the phase 2 Security Association due to the movement of a DVIPA. For example, a host-to-host UDP-encapsulated tunnel mode Security Association protecting specific protocols or ports that was initially initiated from a non-z/OS client might not be able to be renegotiated when the z/OS system assuming control of the DVIPA initiates the negotiation. In this case, the Security Association must be reestablished by peer initiation.

FIPS 140 mode and Sysplex-Wide Security Associations

To enable FIPS 140 mode in a sysplex, you should enable FIPS 140 mode for all the TCP/IP stacks, IKE daemons, and network security services (NSS) daemons on all systems in the sysplex. If the FIPS 140 mode of the distributing TCP/IP stack in the sysplex is different than the FIPS 140 mode of the target TCP/IP stacks, distribution of some of the tunnels across the target TCP/IP stacks might not function as expected.

If the distributor is configured with FIPS 140 support, then all of the tunnels that it activates will adhere to the FIPS 140 restrictions (for example, DES encryption is not allowed). All target TCP/IP stacks can use the distributed tunnels and can process the distributed traffic, regardless of their FIPS 140 mode. The targets that are configured without FIPS 140 support might not adhere to the strict cryptographic module boundaries as defined by FIPS 140, but the Security Association will be successfully activated and used.

If the distributor is not configured with FIPS 140 support, it might activate Security Associations that do not adhere to the FIPS 140 restrictions. The distributor can successfully distribute those associations to target TCP/IP stacks that are not configured in FIPS 140 mode. However, if the distributor distributes those associations to target TCP/IP stacks that are configured in FIPS 140 mode, the data flowing over connections to those stacks is discarded because the associated Security Association or tunnel is not installed.

The distributor and its backup can be configured differently with respect to the FIPS 140 mode. When a backup takes over from a distributor, all the IP security tunnels are renegotiated by the backup. The renegotiated tunnels operate at the FIPS 140 mode that is defined on the backup distributor. This can change the operation of the tunnels as they are distributed by the backup. If the distributor is configured with FIPS 140 support and its backup is not, the backup might activate Security Associations that do not adhere to the FIPS 140 restrictions and distribution of them might fail.

For more information about enabling FIPS 140 mode in a sysplex environment, see "Steps for configuring IP security to support FIPS 140 mode" on page 891.

Shadow Security Associations

When an IP security stack is the target of a DVIPA, it receives a copy (shadow) of any active Security Associations for the DVIPA. To display the shadow Security Associations, use the following command:

```
ipsec -y display -s

CS V1R12 ipsec  Stack Name: TCPCS  Tue Feb 16 10:39:25 2010
Primary:  Dynamic tunnel  Function: display (shadows)  Format:   Detail
Source:   Stack           Scope:    Current            TotAvail: 1

TunnelID:                     Y2
Generation:                   1
IKEVersion:                   1.0
ParentIKETunnelID:            K1
VpnActionName:                TransportMode
LocalDynVpnRule:              n/a
State:                        Active
HowToEncap:                   Transport
LocalEndPoint:                9.1.1.1
RemoteEndPoint:               9.1.1.2
LocalAddressBase:             9.1.1.1
LocalAddressPrefix:           n/a
LocalAddressRange:            n/a
RemoteAddressBase:            9.1.1.2
RemoteAddressPrefix:          n/a
RemoteAddressRange:           n/a
HowToAuth:                    ESP
 AuthAlgorithm:               HMAC-MD5
 AuthInboundSpi:              1878088104 (0x6FF159A8)
 AuthOutboundSpi:             270783814  (0x1023D546)
HowToEncrypt:                 DES-CBC
 KeyLength:                   n/a
 EncryptInboundSpi:           1878088104 (0x6FF159A8)
 EncryptOutboundSpi:          270783814  (0x1023D546)
Protocol:                     ALL(0)
LocalPort:                    n/a
LocalPortRange:               n/a
RemotePort:                   n/a
RemotePortRange:              n/a
Type:                         n/a
TypeRange:                    n/a
Code:                         n/a
CodeRange:                    n/a
OutboundPackets:              1
OutboundBytes:                264
InboundPackets:               1
InboundBytes:                 264
Lifesize:                     0K
LifesizeRefresh:              0K
CurrentByteCount:             0b
LifetimeRefresh:              2010/02/16 14:26:22
LifetimeExpires:              2010/02/16 14:37:43
CurrentTime:                  2010/02/16 10:39:25
VPNLifeExpires:               2010/02/17 10:37:43
NAT Traversal Topology:
  UdpEncapMode:               No
```

```
  LclNATDetected:             No
  RmtNATDetected:             No
  RmtNAPTDetected:            No
  RmtIsGw:                    n/a
  RmtIsZOS:                   n/a
  zOSCanInitP2SA:             n/a
  RmtUdpEncapPort:            n/a
  SrcNATOARcvd:               n/a
  DstNATOARcvd:               n/a
PassthroughDF:                n/a
PassthroughDSCP:              n/a
************************************************************************

1 entries selected
```

Sample IP security policy files

A sample stack-specific policy is located in /usr/lpp/tcpip/samples/pagent_IPSec.conf.

A sample common policy is located in /usr/lpp/tcpip/samples/pagent_CommonIPSec.conf.

Chapter 18. Network security services

A network security services (NSS) server provides network security services for one or more security disciplines. Supported disciplines are IPSec, which is a set of services that supports IPSec and IKE processing, and XMLAppliance, which is a set of services for XML appliances. For the IPSec discipline, these services include the IPSec certificate service and the IPSec remote management service. For the XMLAppliance discipline, the NSS server supports the XMLAppliance SAF access service, the XMLAppliance certificate service, and the XMLAppliance private key service. For information about configuring the IKE daemon to act as an NSS IPSec client on behalf of a TCP/IP stack, see Chapter 17, “IP security,” on page 883.

Terms and concepts for network security services

The following terms and concepts apply to the information about network security services (NSS):

certificate bundle
An X.509 bundle as defined in Section 3.6 of RFC 5996, *Internet Key Exchange Protocol: IKEv2*. A certificate bundle can contain multiple DER encoded certificates and certificate revocation lists (CRLs). You can use the **certbundle** command to create a certificate bundle.

Certificate revocation list (CRL)
A time-stamped list of revoked certificates that is signed by a certificate authority.

CRLDistributionPoints
An optional X.509 certificate extension that identifies one or more locations where the CRL for a certificate is.

hash and URL encoding
A certificate payload encoding that includes the hash of a certificate or bundle and the URL that identifies where that certificate or bundle can be retrieved from an HTTP server

IPSec certificate service
A service for NSS IPSec clients that provides IPSec digital signature and verification services.

IPSec discipline
A set of services provided to an NSS IPSec client. The services are the IPSec certificate service and the IPSec remote management service.

IPSec remote management service
A service for NSS IPSec clients that provides remote IPSec management capability.

Network security services (NSS)
A set of services that performs security enforcement or management. The services are provided in groupings called security disciplines.

NSS client
A client that requests network security services from an NSS server.

NSS daemon (NSSD)
The z/OS UNIX daemon that implements the NSS server functionality.

NSS IPSec client
An NSS client that is using the IPSec discipline. The z/OS IKE daemon can act as an NSS IPSec client for one or more TCP/IP stacks.

NSS server
Provides network security services for one or more NSS clients.

NSS XMLAppliance client
An NSS client that is using the XMLAppliance discipline.

security discipline
A specific grouping of network security services.

© Copyright IBM Corp. 2000, 2022

trust chain
The signing sequence of certificates for any particular certificate back to a root certificate authority.

XML appliance
A network appliance that processes XML messages efficiently and securely. XML appliances often offload XML parsing and transformations from host systems and implement a variety of XML security features.

XMLAppliance certificate service
A service for NSS XMLAppliance clients that provides key ring listing and certificate retrieval capability.

XMLAppliance discipline
A set of services provided to an NSS XMLAppliance client. The NSS server supports the XMLAppliance SAF access service, the XMLAppliance certificate service, and the XMLAppliance private key service.

XMLAppliance private key service
A service for NSS XMLAppliance clients that provides private key retrieval of private keys that are not protected by Integrated Cryptographic Service Facility (ICSF), RSA signature generation using ICSF-protected private keys, and RSA message decryption using ICSF-protected private keys.

XMLAppliance SAF access service
A service for NSS XMLAppliance clients that provides SAF user authentication and access control capability.

For additional IP security-related terms, see Chapter 17, “IP security,” on page 883.

Network security services overview

Network security services (NSS) includes services that perform security enforcement or management. As shown in Figure 136 on page 1084, NSS includes services provided by the NSS IPSec discipline and NSS XMLAppliance discipline. Each discipline includes a subset of services provided by NSS and is intended for use by a specific type of NSS client.

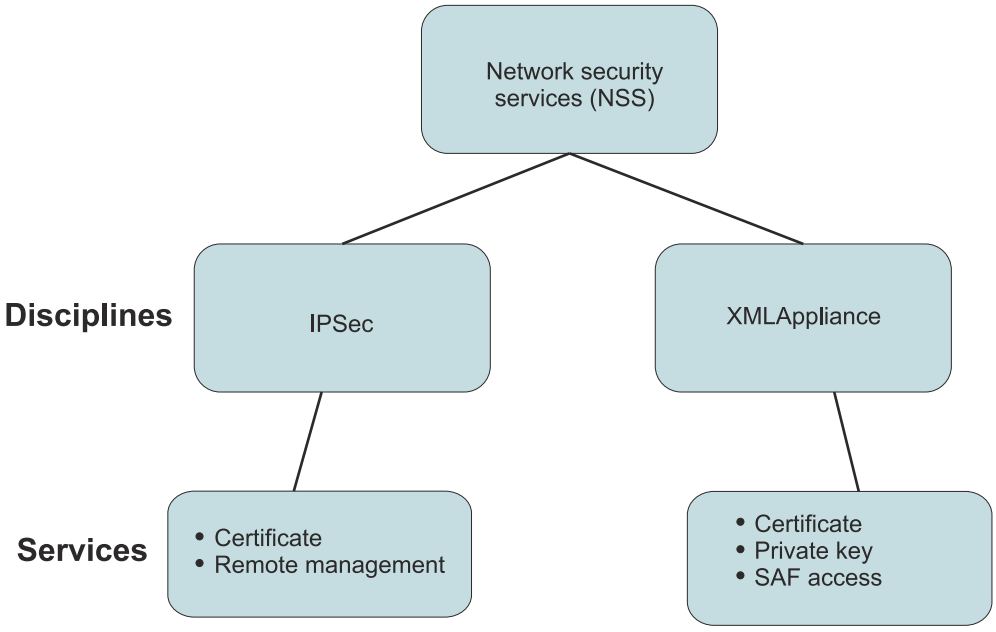


Figure 136. NSS services by discipline

NSS IPSec discipline overview

The NSS IPSec discipline includes the NSS IPSec certificate service and NSS IPSec remote management service.

The NSS server provides the NSS IPSec certificate service to perform digital signature and verification operations on behalf of an NSS IPSec client. The NSS IPSec certificate service is used by an NSS IPSec client during a phase 1 negotiation when digital signature authentication is required. Certificates and private keys for all NSS IPSec clients are stored on a single key ring. The NSS server must have access to this key ring and must have access to the certificates and private keys on this key ring. When providing the NSS IPSec certificate service, the NSS server consults SERVAUTH profiles to verify that an NSS IPSec client is authorized to access the certificates that are involved. For details about these profiles, see step “9.d” on page 1089 and step “9.e” on page 1089, under “Steps for authorizing resources for NSS” on page 1086.

The NSS server uses the NSS IPSec remote management service to request IPSec monitoring data from an NSS IPSec client and to make IPSec control requests to an NSS IPSec client. Control requests include the ability to activate, deactivate, or refresh a Security Association, and to switch between default IP filter policy and IP security filter policy. Use the **ipsec** command and the IPSec network management interface (NMI) to make these requests. For details about the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands. For details about the IPSec NMI, see z/OS Communications Server: IP Programmer's Guide and Reference.

NSS XMLAppliance discipline

The NSS XMLAppliance discipline includes the NSS XMLAppliance SAF access service, the NSS XMLAppliance certificate service, and the NSS XMLAppliance private key service.

An NSS XMLAppliance client uses the NSS XMLAppliance SAF access service to perform SAF user authentication and access control checks. The NSS server consults SERVAUTH profiles for access control checks. For details about these profiles, see step “9.d” on page 1089 and step “9.e” on page 1089, under “Steps for authorizing resources for NSS” on page 1086.

The NSS XMLAppliance certificate service enables an NSS server to provide a list of authorized certificates on its key ring. Those certificates can then be retrieved on behalf of an NSS XMLAppliance client. Certificates for all NSS XMLAppliance clients are stored on a single key ring. The NSS server must have access to this key ring and must have access to the certificates on this key ring. When the NSS server provides the NSS XMLAppliance certificate service, it consults SERVAUTH profiles to verify that an NSS XMLAppliance client is authorized to access the certificates involved. For details about these profiles, see step “9.d” on page 1089 and step “9.e” on page 1089, in “Steps for authorizing resources for NSS” on page 1086.

An NSS XMLAppliance client uses the NSS XMLAppliance private key service to retrieve authorized private keys stored in the SAF database of the NSS server. The private key service also enables the NSS server to perform RSA signature and RSA decryption operations using private keys protected by Integrated Cryptographic Service Facility (ICSF) on behalf of an NSS XMLAppliance client. An NSS XMLAppliance client can use a retrievable private key to sign and decrypt XML messages locally. XML appliances that are in less-trusted network zones can use a centralized NSS server to perform critical RSA operations using ICSF-protected private keys on behalf of the appliance. Certificates and private keys for all NSS XMLAppliance clients are stored on a single key ring. The NSS server must have access to this key ring and must have access to the certificates and private keys on this key ring. Retrieval of the private key is not allowed if the private key is stored in the ICSF public key data set (PKDS). When providing NSS XMLAppliance private key service, the NSS server consults SERVAUTH profiles to verify that an NSS XMLAppliance client is authorized to access the certificates and associated keys involved. For details about these profiles, see step “9.d” on page 1089, step “9.e” on page 1089, and step “9.g” on page 1090 under “Steps for authorizing resources for NSS” on page 1086.

Preparing to provide network security services

Before network security services can be provided, authorization to several resources must be defined to the external security manager. This topic also includes NSS server certificate label naming considerations, an NSS client authorization example, information on configuring and controlling the NSS server, and recovery considerations.

Steps for authorizing resources for NSS

Before network security services can be provided, authorization to several resources must be defined to the external security manager.

Before you begin

RACF is used as the external security manager in the following examples. RACF commands shown in these examples are also provided in the EZARACF member of the SEZAINST data set. In these examples, it is assumed that the NSS server is running under the user ID NSSD.

Procedure

Perform the following steps to authorize access to the appropriate resources:

1. Define and authorize the NSSD user ID.

   The NSS server is a z/OS UNIX application that you can start from the z/OS UNIX shell or from an MVS started procedure. Before starting the NSS server, you must define the NSSD user ID to the external security manager with UID 0. If you start the NSS server from an MVS started procedure, the NSSD user ID must also be authorized to the STARTED class.

   Issue the following commands:

   ```
   ADDUSER  NSSD     DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(0)  HOME('/'))
   RDEFINE  STARTED  NSSD.*            STDATA(USER(NSSD))
   SETROPTS RACLIST(STARTED) REFRESH
   SETROPTS GENERIC(STARTED) REFRESH
   ```

2. Permit the NSSD user ID to SYS1.PARMLIB.

   The NSS server uses the TCP/IP component trace (CTRACE) to perform service-level tracing. The default NSS server component trace parmlib member is stored in SYS1.PARMLIB. The NSSD user ID must be permitted to access SYS1.PARMLIB.

   Issue the following command:

   ```
   PERMIT   SYS1.PARMLIB  ID(NSSD) ACCESS(READ)
   ```

3. Define key ring controls.

   Certificates used by NSS clients are stored on a SAF key ring. The RACDCERT command is used to manage a RACF key ring. The IRR.DIGTCERT FACILITY class resource is used to control access to the RACDCERT command. If these controls do not already exist, they must be defined as follows:

   ```
   RDEFINE FACILITY IRR.DIGTCERT.ADD UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.ADDRING UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.CONNECT UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.GENCERT UACC(NONE)
   RDEFINE FACILITY IRR.DIGTCERT.GENREQ UACC(NONE)
   ```

   For details about these controls, see z/OS Security Server RACF Command Language Reference.

4. Give the user ID of the administrator that will manage the NSS server's key ring appropriate access to manage the key ring.

   Issue the following commands:

   ```
   PERMIT IRR.DIGTCERT.ADD CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.ADDRING CLASS(FACILITY) ID(userid) ACC(UPDATE)
   PERMIT IRR.DIGTCERT.CONNECT CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.GENCERT CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.GENREQ CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.LIST CLASS(FACILITY) ID(userid) ACC(CONTROL)
   PERMIT IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(userid) ACC(UPDATE)
   ```

   After you permit access to the various IRR.DIGTCERT resources, update the FACILITY class as follows:

```
SETROPTS RACLIST(FACILITY) REFRESH
```

5. Give the NSSD user ID access to the key ring

   You must provide appropriate access for the NSSD user ID to its own key ring using the RDATALIB or FACILITY class. The RDATALIB class is preferred as it allows for granular authorization to a specified key ring. The FACILITY class allows global authorization to all key rings and certificates. See z/OS Security Server RACF Callable Services for additional information.

   To provide access using the RDATALIB class:

   ```
   a. Activate and RACLIST the RDATALIB class if not already active:
   SETROPTS CLASSACT(RDATALIB) RACLIST(RDATALIB)

   b. Define an RDATALIB profile for the NSSD key ring. Replace keyringname with the name of
   the configured key ring.
   REDEFINE RDATALIB NSSD.keyringname.LST UACC(NONE)

   c. Permit access to the RDATALIB profile for the NSSD user ID.
   PERMIT NSSD.keyringname.LST CLASS(RDATALIB) ID(NSSD) ACCESS(READ)

   d. When using RDATALIB to control list access to the NSSD key ring, the administrator's ID
   should also be permitted to the RDATALIB profile.
   PERMIT NSSD.keyringname.LST CLASS(RDATALIB) ID(userid) ACC(UPDATE)

   e. Refresh the RDATALIB class
   SETROPTS RACLIST(RDATALIB) REFRESH
   ```

   To provide access using the FACILITY class:

   ```
   PERMIT IRR.DIGTCERT.LIST CLASS(FACILITY) ID(NSSD) ACC(READ)
   PERMIT IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(NSSD) ACC(READ)
   ```

   After you permit access to the various IRR.DIGTCERT resources, update the FACILITY class as follows:

   ```
   SETROPTS RACLIST(FACILITY) REFRESH
   ```

6. Optionally, permit the NSS server to the BPX.DAEMON FACILITY class profile.

   For information concerning the use of the BPX.DAEMON profile, see "BPX.DAEMON FACILITY class profile" on page 37.

   If you decide to use this profile, permit the NSS server user ID to this profile using the following command, where the *userid* value is the user ID under which the NSS server runs:

   ```
   PERMIT BPX.DAEMON CLASS(FACILITY) ID(userid) ACCESS(READ)
   ```

7. Enable the secured signon function.

   The NSS server supports the use of PassTickets. To use this support, the secured signon function must be enabled, and at least one profile must be created for the NSS server.

   The secured signon function of RACF is enabled by activating the PTKTDATA class as follows:

   ```
   SETROPTS CLASSACT(PTKTDATA)
   SETROPTS RACLIST(PTKTDATA) REFRESH
   ```

   Profiles in the PTKTDATA class control the use of the secured signon function by an application. A secured signon application key is associated with each profile. This key is stored in the external security manager's database. When RACF is used as the external security manager, this key is stored in a masked or encrypted state.

   With RACF, you can use a secured signon application key that is controlled at the following levels:

   - All users who need access to the application
   - A specific RACF group of users who need access to the application
   - A specific RACF user, when connected to a specific RACF group

- A specific RACF user

The application name NSSD must be used when defining the secured signon keys for the NSS server. The following example shows a RACF command that you can issue to assign a secured signon key that can be used by all NSS clients that are authenticating to the NSS server:

```
RDEFINE PTKTDATA NSSD SSIGNON(KEYMASKED(E001193519561977)) UACC(NONE)
```

For specific information about enabling the secured signon function and defining profiles to be used by the single signon function, see z/OS Security Server RACF Security Administrator's Guide.

8. Define SERVAUTH profiles to authorize NSS clients to network security services.

These profiles are on the same system as the NSS server. Many of these profiles are constructed using the name of an NSS client. NSS clients must authenticate to the NSS server using a valid user ID and password, or a valid user ID and PassTicket. This user ID must be given access to the SERVAUTH profiles created on behalf of the NSS client.

For details about how to define the IKE daemon as an NSS client, see "Using network security services" on page 1042.

Perform the following steps to authorize NSS clients:

a. Define a SAF user ID representing an NSS client to the external security manager.

An NSS client must present valid credentials to the NSS server before accessing any services. Valid credentials include a user ID and password, or a user ID and PassTicket if secured signon is enabled. A SAF user ID representing an NSS client must be defined to the external security manager.

Issue the following command:

```
ADDUSER  userid   DFLTGRP(OMVSGRP)  OMVS(UID(x))
```

Rules:

- Multiple NSS clients can use a single user ID. However, each NSS client must have a unique client name.
- A SAF user ID must have an OMVS segment with either AUTOID or a specific UID(*x*) defined for the NSSD to authenticate it as an NSS client.

Guideline: Because SAF user IDs are used to authorize a client to the NSS services, avoid sharing a single user ID across NSS disciplines.

b. If you choose to define an NSSD profile in the APPL class with UACC(NONE), issue the following command to authorize each SAF user ID to the NSSD application:

```
PERMIT NSSD CLASS(APPL) ID(userid) ACC(READ)
SETROPTS RACLIST(APPL) REFRESH
```

c. Authorize the user ID associated with an NSS client for each of the network security services it will use.

To authorize an NSS client to use a network security service, you must create a SERVAUTH resource profile for that service that represents the NSS client. The user ID associated with the NSS client must be permitted READ access to that profile. Table 54 on page 1088 shows the name of the SERVAUTH profile for each service, where *sysname* is the name of the z/OS system running the NSS server and *clientname* is the name by which the NSS server knows the NSS client.

Table 54. SERVAUTH profile names for NSS

Service	SERVAUTH profile name
IPSec certificate service	EZB.NSS.*sysname*.*clientname*.IPSEC.CERT

Table 54. SERVAUTH profile names for NSS (continued)	
Service	**SERVAUTH profile name**
IPSec remote management service	EZB.NSS.*sysname.clientname*.IPSEC.NETMGMT
XMLAppliance certificate service	EZB.NSS.*sysname.clientname*.XMLAPPLIANCE.CERT
XMLAppliance private key service	EZB.NSS.*sysname.clientname*.XMLAPPLIANCE.PRIVKEY
XMLAppliance SAF access service	EZB.NSS.*sysname.clientname*.XMLAPPLIANCE.SAFACCESS

You can authorize the NSS client to a SERVAUTH profile using the following commands:

```
RDEFINE SERVAUTH profile_name  UACC(NONE)
PERMIT  profile_name (SERVAUTH) ID(nssclient) ACCESS(READ)
SETROPTS GENERIC(SERVAUTH) REFRESH
SETROPTS RACLIST(SERVAUTH) REFRESH
```

Tip: You can use a wildcard in the profiles to reduce the number of profile entries that must be defined.

d. Create a SERVAUTH resource profile for each NSS IPSec client certificate added to the NSS server's key ring, and give each NSS IPSec client's user ID access to the profiles created for its own certificates.

The name of such a resource profile is EZB.NSSCERT.*sysname.mappedlabelname*.HOST, where *sysname* is the name of the z/OS system running the NSS server and *mappedlabelname* is the mapped name of the certificate's label in the key ring. For details about determining a certificate label's mapped name, see “NSS server certificate label naming considerations” on page 1092.

This can be accomplished with the following commands:

```
RDEFINE SERVAUTH EZB.NSSCERT.sysname.mappedlabelname.HOST UACC(NONE)
PERMIT EZB.NSSCERT.sysname.mappedlabelname.HOST CLASS(SERVAUTH) ID(userid) ACCESS(READ)
SETROPTS GENERIC(SERVAUTH) REFRESH
SETROPTS RACLIST(SERVAUTH) REFRESH
```

e. Create a SERVAUTH resource profile for each certificate authority (CA) certificate that could be used by an NSS IPSec client, and give the NSS client's user ID access to the profiles.

When the digital signature mode of authentication is used, an NSS IPSec client can provide a remote security endpoint with information about certificate authorities that are trusted by the NSS client. The remote security endpoint should use this information as a hint to decide which of its certificates to use when creating its signature.

By default, an NSS IPSec client sends a remote security endpoint information about all the certificate authorities that the NSS IPSec client is authorized to advertise. This can result in an NSS IPSec client sending a large amount of data to a remote security endpoint. Use the CaLabel parameter on the RemoteSecurityEndpoint statement to reduce the amount of data sent to specific remote security endpoints. For details about the RemoteSecurityEndpoint statement, see z/OS Communications Server: IP Configuration Reference.

A SERVAUTH resource profile is used to authorize NSS IPSec clients to use a CA. A SERVAUTH resource profile must be created for each CA certificate that could be used by an NSS IPSec client. The name of such a resource profile is EZB.NSSCERT.*sysname.mappedlabelname*.CERTAUTH, where *sysname* is the name of the z/OS system running the NSS server and *mappedlabelname* is the mapped name of the certificate's label in the key ring. For details about determining a certificate label's mapped name, see “NSS server certificate label naming considerations” on page 1092. An NSS IPSec client's user ID must be given access to this profile before it can use the corresponding CA certificate.

This can be accomplished with the following commands:

```
RDEFINE  SERVAUTH EZB.NSSCERT.sysname.mappedlabelname.CERTAUTH UACC(NONE)
PERMIT   EZB.NSSCERT.sysname.mappedlabelname.CERTAUTH CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
SETROPTS GENERIC(SERVAUTH) REFRESH
SETROPTS RACLIST(SERVAUTH) REFRESH
```

f. Create a SERVAUTH resource profile for each certificate that an NSS XMLAppliance client could retrieve and give the user ID of the NSS XMLAppliance client access to the appropriate profiles.

   In contrast to the IPSec discipline, the XMLAppliance discipline does not distinguish between host, site, or certificate authority certificates. For any given certificate request, the NSS server first checks the EZB.NSSCERT.*sysname.mappedlabelname*.HOST profile. If that check fails, the server then checks the EZB.NSSCERT.*sysname.mappedlabelname*.CERTAUTH profile. If the NSS XMLAppliance client has read access to either profile, the server permits access to the certificate resource. It is up to the NSS server administrator to allow or deny access to each certificate, and it is up to the XML appliance administrator to determine how each certificate should be used. The user ID of an NSS XMLAppliance client must be given access to one of these profiles before it can use the corresponding certificate. For examples on how to define the required profiles and permit access, see step "9.d" on page 1089 and step "9.e" on page 1089, under "Steps for authorizing resources for NSS" on page 1086. For details about determining the mapped name of a certificate label, see "NSS server certificate label naming considerations" on page 1092.

g. Create a SERVAUTH resource profile for the private key of each certificate to which an NSS XMLAppliance client requires access and give the user ID of the NSS XMLAppliance client access to the profiles.

   Use the SERVAUTH resource profile to authorize NSS XMLAppliance clients to retrieve the private key from a certificate, and to locally perform any RSA operations that are based on the private key or to make ICSF calls requesting RSA operations on System z for ICSF-protected keys. You must create a SERVAUTH resource profile for each private key against which the XMLAppliance client needs to perform operations. The name of such a resource profile is EZB.NSSCERT.*sysname.mappedlabelname*.PRIVKEY, where *sysname* is the name of the z/OS system running the NSS server and *mappedlabelname* is the mapped name of the certificate label in the key ring. Ensure that the user ID of an NSS XMLAppliance client has appropriate access to this profile so that it can retrieve the private key or perform any RSA operations that are based on the private key. This can be accomplished with the following commands:

```
RDEFINE  SERVAUTH EZB.NSSCERT.sysname.mappedlabelname.PRIVKEY UACC(NONE)
PERMIT   EZB.NSSCERT.sysname.mappedlabelname.PRIVKEY CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
SETROPTS GENERIC(SERVAUTH) REFRESH
SETROPTS RACLIST(SERVAUTH) REFRESH
```

   For details about determining the mapped name of a certificate label, see "NSS server certificate label naming considerations" on page 1092.

h. Create the following SERVAUTH profiles to enable users to remotely monitor (IPSEC.DISPLAY) or manage (IPSEC.CONTROL) NSS clients:

   - EZB.NETMGMT.*sysname.clientname*.IPSEC.DISPLAY
   - EZB.NETMGMT.*sysname.clientname*.IPSEC.CONTROL

   For more information, see the information about the `-z` option of the **ipsec** command in "NSS client authorization example" on page 1093. For more details about managing network security, see z/OS Communications Server: IP System Administrator's Commands. For details about the network management interface, see z/OS Communications Server: IP Programmer's Guide and Reference.

9. If you are using the NSS XMLAppliance private key service with ICSF-protected private keys, authorize the NSS server to the Integrated Cryptographic Service Facility (ICSF).

   ICSF is required when using the RSA operations in the XMLAppliance private key service. The XMLAppliance private key service uses ICSF in the following ways:

- Encrypting signature data for the XMLAppliance private key service RSA signature generation message flow.
- Decrypting data for the XMLAppliance private key service RSA decryption message flow.

ICSF provides cryptography support through various cryptographic hardware features. The cryptographic features that are available to your applications depend on your processor or server model. For information about which features are available on your hardware, see the information about callable service support by hardware configuration in z/OS Cryptographic Services ICSF Overview. For details about configuring ICSF, see z/OS Cryptographic Services ICSF Administrator's Guide.

When using a cryptographic coprocessor, the callable ICSF service names that are used by the XMLAppliance certificate service are as follows:

- CSNDDSG
- CSNDPKD

Requirement: If you plan to use the RSA operations within the XMLAppliance private key service, the NSS server must be permitted to access the ICSF cryptographic services (CSFSERV). Use the following commands to define the appropriate profiles in the CSFSERV class, give the NSS server access to the profiles, activate the CSFSERV class, and refresh the RACF profiles in storage:

```
RDEFINE service-name CLASS(CSFSERV) UACC(NONE)
PERMIT service-name CLASS(CSFSERV) ID(server-name) ACCESS(READ)
SETROPTS CLASSACT(CSFSERV)
SETROPTS RACLIST(CSFSERV) REFRESH
```

10. If XMLAppliance clients using the SAF access service are using certificates for access checks, enable RACF certificate name filtering.

    The NSS XMLAppliance SAF access service can use RACF certificate name filtering to map an X.500 distinguished name to a RACF ID when performing SAF access checks. The DIGTNMAP class must be active to perform certificate name filtering. Activate the DIGTNMAP class with the following commands:

    ```
    SETOPS CLASSACT(DIGTNMAP)
    SETROPTS RACLIST(DIGTNMAP) REFRESH
    ```

    Create a certificate name filter for each mapping of an X.500 distinguished name to a RACF ID using the following commands:

    ```
    RACDCERT ID(userid) MAP SDNFILTER('x500dn')
    SETROPTS RACLIST(DIGTNMAP) REFRESH
    ```

    For specific details on enabling RACF certificate name filtering, see z/OS Security Server RACF Security Administrator's Guide.
11. The NSSD uses ICSF callable services for ECDSA digital signature support. The services it uses are the PKCS11 private key sign service and the PKCS11 public key verify service. You can control access to these services with RACF, using the CSFSERV general resource class, and the CSF1PKS and CSF1PKV profiles.

    If the CSFSERV class is defined, and the CSF1PKS and CSF1PKV profiles are defined, grant the NSSD user ID read access to the defined profiles using the following commands:

    ```
    PERMIT CSF1PKS CLASS(CSFSERV) ID(NSSD) ACCESS(READ)
    PERMIT CSF1PKV CLASS(CSFSERV) ID(NSSD) ACCESS(READ)
    SETROPTS RACLIST(CSFSERV) REFRESH
    ```

    See z/OS Cryptographic Services ICSF Administrator's Guide for more information about the CSFSERV general resource.

NSS server certificate label naming considerations

During the processing of certificate operations, the NSS server validates that an NSS client is authorized to access the certificates required to complete the operation. The NSS server consults SERVAUTH profiles to perform this validation. The profile names consulted by the NSS server are dynamically constructed by the NSS server using the following information:

- The system name on which the NSS server is running
- The label of the certificate this is used during a certificate operation
- The certificate operation that is being performed:
  - When processing a request to create a signature, the format of the profile that is consulted is EZB.NSSCERT.*sysname*.*mappedlabelname*.HOST.
  - When processing a request to obtain a list of CA certificates, the format of the profile consulted is EZB.NSSCERT.*sysname*.*mappedlabelname*.CERTAUTH.
  - When processing a request to retrieve a private key that is not protected by Integrated Cryptographic Service Facility (ICSF) or to use an ICSF-protected private key, the format of the profile consulted is EZB.NSSCERT.*sysname*.*mappedlabelname*.PRIVKEY.

The NSS server creates a mapped label name using the following algorithm:

- All lowercase alphabetic characters in a certificate's label are changed to uppercase. This is necessary because the class descriptor table for the SERVAUTH profile permits only uppercase profile names.
- The asterisk (*), percent sign (%), and ampersand (&) are replaced by a dollar sign ($). This is necessary because these characters have special meaning when generic profile processing is active.
- All embedded blanks are also replaced by a dollar sign ($). This is necessary because blanks are not allowed in SERVAUTH profile names.

Rules:

- The administrator of the NSS server must define profiles using the mapped label names generated by this algorithm. When the certificate's label name contains lowercase characters, the administrator must change each lowercase character to uppercase. When the certificate's label name contains the characters *, %, &, or a blank character, the administrator must replace each occurrence with a dollar sign ($) character.
- When a certificate label contains the period character (.), ensure that the corresponding SERVAUTH profile contains matching qualifiers. For example, if you request a certificate with the label CERTIFICATE.123.ABC for a private key operation, the NSS server checks a SERVAUTH profile named EZB.NSSCERT.*sysname*.CERTIFICATE.123.ABC.PRIVKEY; defining a SERVAUTH profile named EZB.NSSCERT.*sysname*.CERTIFICATE.*.PRIVKEY does not permit access to the private key of the certificate.

Using this algorithm, it is possible that multiple certificates can result in the same mapped name. This is shown in Table 55 on page 1092.

Table 55. Mapped label names

Label	**Mapped label**
CERTIFICATE_123	CERTIFICATE_123
Certificate_123	CERTIFICATE_123
CERTIFICATE 123	CERTIFICATE$123
CERTIFICATE%123	CERTIFICATE$123
CERTIFICATE*123	CERTIFICATE$123
CERTIFICATE&123	CERTIFICATE$123
CERTIFICATE$123	CERTIFICATE$123

Tip: When creating certificates for the NSS server's key ring, avoid using lowercase alphabetic characters, blanks, and the characters *, %, and & in the certificate's label.

NSS client authorization example

Consider the configuration shown in Figure 137 on page 1093.

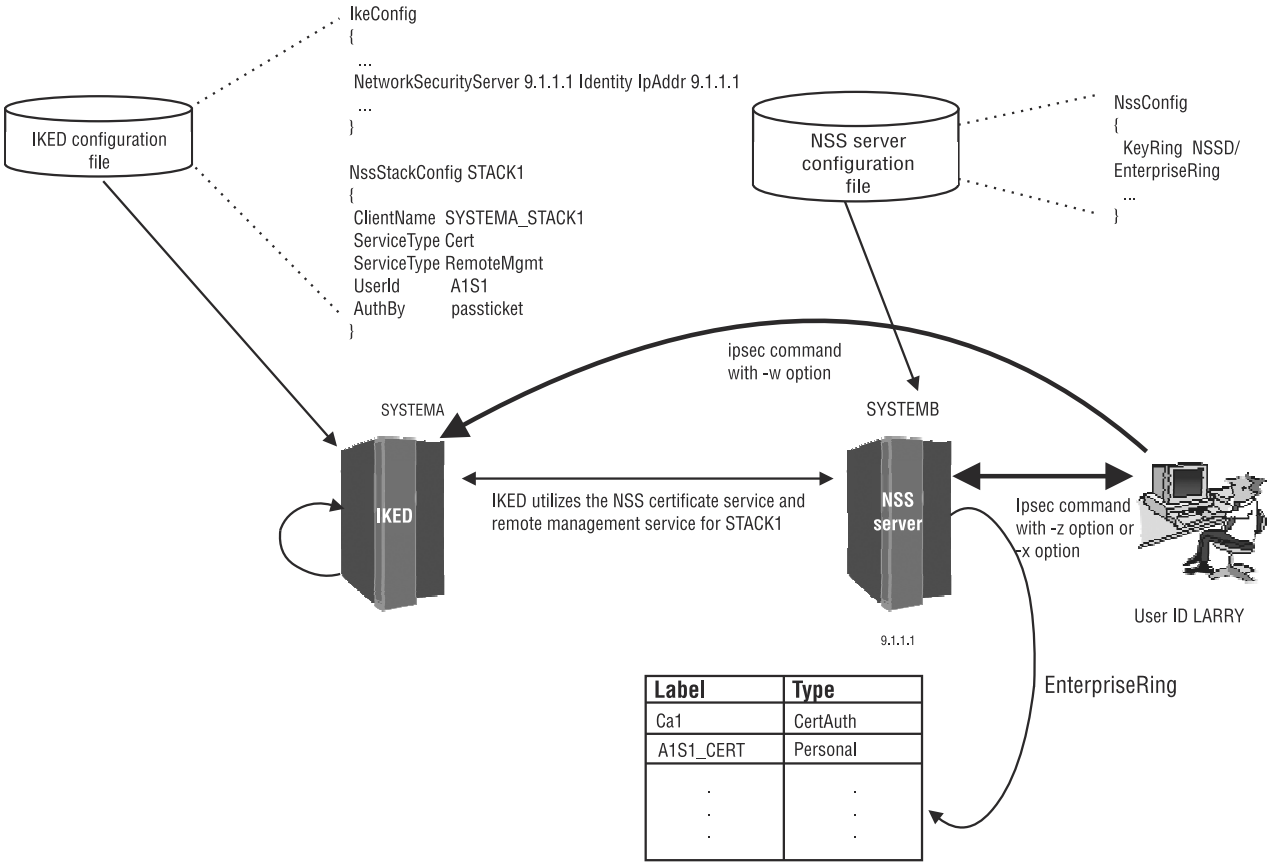


Figure 137. NSS client authorization example

In this example, note the following configuration:

- Stack STACK1 is defined as an NSS IPSec client in the IKE daemon configuration file on system SYSTEMA. The client name for STACK1 is SYSTEMA_STACK1. The user ID associated with SYSTEMA_STACK1 is A1S1. The user ID A1S1 must be defined to the external security manager on system SYSTEMB.
- Client SYSTEMA_STACK1 is configured to use the NSS certificate service. On system SYSTEMB, the user ID A1S1 must be given read access to the following SERVAUTH profile:

```
EZB.NSS.SYSTEMB.SYSTEMA_STACK1.IPSEC.CERT
```

- Client SYSTEMA_STACK1 is configured to use the NSS remote management service. On system SYSTEMB, the user ID A1S1 must be given read access to the following SERVAUTH profile:

```
EZB.NSS.SYSTEMB.SYSTEMA_STACK1.IPSEC.NETMGMT
```

- The NSS server on system SYSTEMB is configured to use the key ring EnterpriseRing. This key ring is owned by the NSSD user ID. Certificates for all NSS IPSec clients are stored on this key ring.

  The NSS server's AT-TLS policy must also specify a key ring from which to obtain the NSS server's personal certificate for use during the TLS negotiation with an NSS client. The NSS server's AT-TLS policy can specify the same key ring as the NSS server's configuration file, or it can specify a different key ring. In either case, the AT-TLS policy should specify which personal certificate to use to represent the NSS server by using the CertificateLabel parameter on the TTLSConnectionAdvancedParms

statement. If this parameter is not configured, AT-TLS attempts to use the default certificate, if one exists, on the configured key ring. If no default certificate exists on the configured key ring and the CertificateLabel parameter is not configured, the TLS negotiation between the NSS client and the NSS server will fail.

The IKE daemon's AT-TLS policy also specifies a key ring. This key ring is used to locate the certificate that was used to sign the NSS server's personal certificate. If the IKE deamon's AT-TLS key ring does not contain this signing certificate, TLS negotiation will fail to verify the NSS server's certificate and the TLS negotiation between the NSS client and the NSS server will fail.

In this example, there is one Personal certificate stored on the key ring for client SYSTEMA_STACK1. On system SYSYEMB, the user ID A1S1 must be given read access to the following SERVAUTH profile before the NSS server can use this certificate to create a signature for client SYSTEMA_STACK1:

```
EZB.NSSCERT.SYSTEMB.A1S1_CERT.HOST
```

In this example, there is also one CertAuth certificate stored on the key ring that should be advertised to IPSec peers by client SYSTEMA_STACK1. On system SYSYEMB, the user ID A1S1 must be given read access to the following SERVAUTH profile before the NSS server can inform client SYSTEMA_STACK1 that it can advertise this CERTAUTH certificate to its peers:

```
EZB.NSSCERT.SYSTEMB.CA.CERTAUTH
```

- The user LARRY will issue the **ipsec** command with the `-z` option to monitor and manage client SYSTEMA_STACK1. On system SYSTEMB, the user ID LARRY must be given read access to the following SERVAUTH profiles:

  ```
  EZB.NETMGMT.SYSTEMB.SYSTEMA_STACK1.IPSEC.DISPLAY
  EZB.NETMGMT.SYSTEMB.SYSTEMA_STACK1.IPSEC.CONTROL
  ```

  The user LARRY will also issue the **ipsec** command with the `-x` option to display information about the NSS server. On system SYSTEMB, the user ID LARRY must be given read access to the following SERVAUTH profile:

  ```
  EZB.NETMGMT.SYSTEMB.SYSTEMB.NSS.DISPLAY
  ```

  In addition, the user LARRY will issue the **ipsec** command with the `-w` option to display information from the IKE daemon about NSS IPSec clients. On system SYSTEMA, the user ID LARRY must be given read access to the following SERVAUTH profile:

  ```
  EZB.NETMGMT.SYSTEMA.SYSTEMA.IKED.DISPLAY
  ```

  Tip: A wildcard can be used in the profiles to reduce the number of profile entries that must be defined.

NSS server configuration considerations

This topic describes configuration issues specific to the NSS server.

Run-time environment

The NSS server is a z/OS UNIX application; it requires the z/OS UNIX file system. The NSS server can be started from an MVS started procedure, from the z/OS shell, with the AUTOLOG statement in the TCP/IP profile, or by using the COMMND*xx* member of PARMLIB. The NSS server must be started by a RACF-authorized user ID, and it must be in an APF-authorized library. For more information about how to start the NSS server, see “Starting the NSS server” on page 1102.

The NSS server uses the MVS operator's console, syslogd, CTRACE, and STDOUT for its logging and tracing. The MVS operator's console and STDOUT are used for major events such as initialization, termination, and error conditions. Syslogd is used for logging events related to the processing of NSS requests. CTRACE is used for detailed tracing and debugging.

The NSS server uses a standard message catalog. The message catalog must be in the UNIX file system. The directory location for the message catalog path is set by the environment variables NLSPATH and LANG.

The NSS server uses ICSF and System SSL for encryption and key management services to provide certificate services to NSS IPSec clients. If the NSS IPSec clients are configured in FIPS 140 mode, you must also configure the NSS server in FIPS 140 mode so that it invokes ICSF and System SSL in FIPS 140 mode. This configuration is required for the entire system to be in FIPS 140 mode.

Language Environment run-time considerations

When starting the NSS server from a started or cataloged procedure, you should typically start it directly from the SEZALOAD data set using PGM=EZANSSD. However, there is a situation in which you might want to start the NSS server using BPXBATCH.

When the NSS server is started using PGM=EZANSSD, the STDENV DD card, if used, is passed directly to the NSS server program. Language Environment does not get access to the STDENV environment variables. As a result, any Language Environment run-time options set in the STDENV DD data set using the _CEE_RUNOPTS environment variable are ignored. In this case, Language Environment run-time options must be passed on the PARM parameter, and the options must be specified before any NSS server options. However, the PARM parameter allows a maximum of 100 characters. If the Language Environment run-time options plus NSS server parameters that you want exceed 100 characters, consider using BPXBATCH to start the NSS server. When PGM=BPXBATCH is used, the Language Environment variable _CEE_RUNOPTS can be included on the STDENV DD card to specify run-time options in excess of 100 characters long.

Steps for configuring the NSS server

The network security services (NSS) server supports the IPSec and XMLAppliance disciplines. IPSec discipline services include the IPSec certificate service and the IPSec remote management service, and XMLAppliance discipline services include the XMLAppliance SAF access service, the XMLAppliance certificate service, and the XMLAppliance private key service.

Procedure

Perform the following steps to configure the NSS server:

1. Create the NSS server configuration file.

   Use the IBM Configuration Assistant for z/OS Communications Server to establish NSS server settings. Establish the settings using the NSS perspective of the IBM Configuration Assistant for z/OS Communications Server, and then use the **Install Configuration File** button on the **Image Information** tab to store the generated NSS server configuration file on the z/OS system.

   Tip: A sample configuration file is provided in /usr/lpp/tcpip/samples/nssd.conf.

   The following search order is used by the NSS server to locate the configuration data set or file:

   a. If the environment variable NSSD_FILE has been defined, the NSS server uses the value as the name of an MVS data set or z/OS UNIX file to access the configuration data.

   b. /etc/security/nssd.conf

   You can specify statements in the configuration file using a variety of EBCDIC code pages. Use the NSSD_CODEPAGE environment variable to specify the code page that you want to use. The default code page is IBM-1047.

   The NSS server configuration file allows the URL of a certificate or certificate bundle that is on an HTTP web server to be associated with the label of a certificate on the key ring of the network security server. See “Using hash and URL certificate encoding types” on page 1099 for additional details.

2. Optionally, set the _BPX_JOBNAME environment variable.

   When starting the NSS server from the z/OS shell, you should set the environment variable _BPX_JOBNAME. This enables a specific job name to be used when reserving ports for the NSS server.

This name can also be used with the STOP or MODIFY console commands. For more information about _BPX_JOBNAME, see z/OS UNIX System Services Planning.

3. Authorize the NSS server to the external security manager, as described in "Steps for authorizing resources for NSS" on page 1086.
4. Configure and start syslogd.

   The NSS server uses the local4 facility when writing messages to syslogd. For performance purposes, syslogd should use z/OS File System as its underlying file system. For more information about syslogd, see "Configuring the syslog daemon" on page 213.
5. Optionally, update the NSS server environment variables.

   The following environment variables are used by the NSS server and can be tailored to a particular installation.

   NSSD_CODEPAGE
   Use the NSSD_CODEPAGE variable to specify the EBCDIC code page to be used when reading the configuration file. For more information about NSSD environment variables and the supported code pages, see z/OS Communications Server: IP Configuration Reference.

   NSSD_CTRACE_MEMBER
   Used by the NSS server to locate a parmlib member for NSS server CTRACE customization. For more information about the TCP/IP services component trace for the NSS server, see z/OS Communications Server: IP Diagnosis Guide.

   NSSD_FILE
   Used by the NSS server in the search order for the NSS server configuration file. For details about the search order used for locating this configuration file, see step "1" on page 1095.

   NSSD_PIDFILE
   Used by the NSS server in the search order for the NSS server PID file. The search order for the NSS server PID file is as follows:

   a. NSSD_PIDFILE environment variable
   b. /etc/nssd.pid
6. Set up the NSS server key ring.

   The NSS server's key ring serves a similar purpose as the IKE daemon's key ring. It contains certificates that are used in the process of creating and verifying signatures that are exchanged during digital signature authentication. A personal certificate or site certificate contained on the key ring of the NSS server represents the identity of an NSS IPSec client, whereas a certificate contained on the IKE daemon's key ring represents a local stack's identity. Certificates for all NSS IPSec clients must be on this one key ring.

   If a personal certificate or site certificate that is contained on the key ring of the NSS server is signed by a certificate authority, then the certificate of that certificate authority must also be connected to the key ring of the NSS server. If the certificate authority is a subordinate certificate authority (such as one that was created by another certificate authority) you should ensure that all the certificate authority certificates that make up the trust chain are connected to the key ring of the NSS server.

   The same commands that are used to create and manage the IKE daemon's key ring also apply to the NSS server's key ring. For examples of how to create and manage the IKE daemon's key ring, see Appendix E, "Steps for preparing to run IP security," on page 1361.

   You must create a SERVAUTH resource profile for each NSS IPSec client certificate that is added to the key ring of the NSS server. For details, see step "9.d" on page 1089.
7. Update the TCP/IP profile and policy files.

   You should update the TCP/IP profile to reserve the port on which the NSS server will listen. If IP security is enabled, consider updating the default IP filter rules in the TCP/IP profile to enable the NSS server to communicate with NSS clients. The IP security policy defined in Policy Agent configuration files must be updated to enable the NSS server to communicate with NSS clients. AT-TLS should be enabled and rules should be defined to protect NSS server communication with NSS clients.

   For additional details concerning these tasks, see "TCP/IP stack considerations" on page 1097.

8. Update the NSS server cataloged procedure (if starting as a started procedure).

   If the NSS server is to be started by a procedure, create the cataloged procedure by copying the sample in SEZAINST(NSSD) to your system. Specify NSS server parameters and change the data set names to suit your local configuration. For a copy of the sample, see z/OS Communications Server: IP Configuration Reference.

Results

If these steps are completed successfully, you should be able to start the NSS server. For details, see "Starting the NSS server" on page 1102.

TCP/IP stack considerations

This topic describes TCP/IP stack considerations, including port reservation, IP filtering, and AT-TLS policy.

Port reservation

By default the NSS server uses TCP port 4159, but this value is configurable using the Port parameter of the NssConfig statement in the NSS server configuration file. For additional details about the NssConfig statement, see z/OS Communications Server: IP Configuration Reference.

Tip: Update the PORT statement in the TCP/IP profile to reserve the port that the NSS server will use when listening for client connections.

```
PORT
     4159 TCP NSSD
```

IP filtering

The NSS server communicates with NSS clients using the TCP protocol. The NSS server binds to all stacks using either INADDR_ANY or in6addr_any as the IP address. IP filters rules must be defined for any IP security stacks that contain an interface to which the NSS client will connect (for details about configuring the IKE daemon as an NSS client, see Chapter 17, "IP security," on page 883). Remote IPSec clients use an ephemeral port when connecting to the NSS server. Ephemeral ports are generally in the range 1024–65355.

Two types of IP filter policy can be defined for a z/OS stack:

- You can define a default IP filter policy in the TCP/IP profile. Updating default IP filter policy to permit communications between the NSS server and NSS clients is optional. Default IP filter policy is in effect only when IP security filter policy cannot be loaded or when the **ipsec -f default** command has been issued.

  For details about defining default IP filter policy in the TCP/IP profile, see z/OS Communications Server: IP Configuration Reference.

  The following default policy contains IPSECRule definitions that allow IPv4 and IPv6 NSS server traffic with NSS clients:

```
IPSEC LOGENable
; Rule      SrcAddr DstAddr   Logging Protocol   SrcPort    DestPort     Routing Secclass

; OSPF protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO OSPF

; IGMP protocol used by Omproute
  IPSECRule *       *         NOLOG   PROTO 2

; DNS queries to UDP port 53
  IPSECRule *       *         NOLOG   PROTO UDP  SRCPort *  DESTport 53

; Administrative access
  IPSECRule *       9.1.1.2   LOG                                          SECCLASS 100

; Network security services (NSS) server access to the NSS client
  IPSECRule *       *         LOG     TCP        SRCPort 4159 DESTport *
```

```
; Network security services (NSS) server access to the NSS client
  IPSEC6Rule *      *         LOG     TCP        SRCPort 4159 DESTport *

ENDIPSEC
```

Rule: The SRCport value in the filter rules must include the value specified on the port parameter of the NssConfig statement in the NSS server configuration file.

- You can define an IP security filter policy in Policy Agent configuration files. IP security filter policy must be updated to permit communications between the NSS server and NSS clients.

  For details about defining IP security policy files, see the Policy Agent and policy applications topic in z/OS Communications Server: IP Configuration Reference.

  An example of an IpFilterRule statement for IPv4, an IpFilterRule statement for IPv6, and an IpGenericFilterAction statement that allows NSS clients to communicate with the NSS server is as follows:

```
IpFilterRule             NssTrafficIPv4
  {
      IpSourceAddr              all4
      IpDestAddrSet             all4
      IpService
      {
         SourcePortRange         4159
         DestinationPortRange    1024 65535
         Protocol                tcp
         Direction               bidirectional InboundConnect
         Routing                 local
      }
      IpGenericFilterActionRef  permit-nolog
  }

IpFilterRule             NssTrafficIPv6
  {
     IpSourceAddr             all6
     IpDestAddrSet            all6
     IpService
     {
        SourcePortRange        4159
        DestinationPortRange   1024 65535
        Protocol               tcp
        Direction              bidirectional InboundConnect
        Routing                local
     }
     IpGenericFilterActionRef  permit-nolog
  }

IpGenericFilterAction    permit-nolog
  {
     IpFilterAction           permit
     IpFilterLogging          no
  }
```

Rule: The DestinationPortRange value on the IpService statements must include the value specified on the port parameter of the NssConfig statement in the NSS server configuration file.

AT-TLS policy

Communications between the NSS server and NSS clients must be secured using Application Transparent Transport Layer Security (AT-TLS). You must define AT-TLS rules to secure this communication. Enable AT-TLS processing for a stack by specifying the TTLS parameter on the TCPCONFIG statement in the TCP/IP profile. Specific AT-TLS policy is configured in Policy Agent configuration files. For details about enabling AT-TLS and configuring AT-TLS policy, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121.

Tip: Define AT-TLS policy such that only cipher suites requiring TLS encryption are exchanged with NSS clients. Failure to restrict the cipher suites to those requiring encryption can result in sensitive information flowing in the clear across an untrusted network.

Rule: You must define AT-TLS policy for each stack through which the NSS server will communicate with an NSS client.

Requirement: The NSS server acts as the server during an SSL handshake. To act in the server role of an SSL handshake, the NSS server must have access to a private key and certificate verifying its ownership of that private key. For information about creating and managing keys and certificates for servers using AT-TLS, see Appendix B, "TLS/SSL security," on page 1329.

A sample AT-TLS policy is located in /usr/lpp/tcpip/samples/pagent_TTLS.conf.

Rule: The LocalPortRange value on the TTLSRule statement must include the value specified on the port parameter of the NssConfig statement in the NSS server configuration file.

Using hash and URL certificate encoding types

During an IKE flow, security endpoints can authenticate each other by exchanging certificate information in certificate payloads and by performing digital signature operations on that certificate information. When the IKED negotiates an IKEv2 phase 1 Security Association on behalf of a network security client, the IKED uses the network security server for these digital signature operations. Encoded certificate information that is received from a remote security endpoint is forwarded to the network security server. Encoded certificate information that is sent to a remote security endpoint by the IKED is obtained from the network security server.

There are several encoding types defined for use in IKEv1, but the only encoding type that must be supported by an IKEv1 implementation is an X.509 certificate signature. Two additional certificate encoding types must be supported by an IKEv2 implementation:

- Hash and URL of an X.509 certificate
- Hash and URL of an X.509 bundle

Generally a hash and URL of a certificate or certificate bundle is considerably smaller than the certificate or the bundle that it represents. Less data is exchanged between IKE peers and between the IKED and the network security server. Although smaller message sizes might improve the efficiency of network resources, using hash and URL encoding requires more processing. The NSSD must communicate with an HTTP server to retrieve the certificate using the URL, and it must validate the certificate using the hash. This communication, when needed, increases the amount of time it takes to complete an IKEv2 phase 1 negotiation.

If you want the NSSD to create hash and URL certificates to send to peers, you must create files on an HTTP server that contain the certificates and the certificate bundles. You must also include the URLs of those files in the NSSD configuration file. See "Enabling the NSSD to generate hash and URL certificate encoding" on page 1099.

Rule: Even if you put a certificate or a certificate bundle on an HTTP server and add the URLs to the NSSD configuration file, you still need to store that certificate on the key ring of the NSSD.

The NSSD can accept hash and URL encoded certificates that it receives from its peers. In that case, the NSSD uses the URLs sent by the peers to locate the certificates. The NSSD caches data that it retrieves from an HTTP server using a URL. See "Enabling the NSSD to process received hash and URL certificate encoding" on page 1100.

Even when URL information is configured for the NSSD, it is ultimately the responsibility of the network security client to decide whether the network security server should use hash and URL encoding. See "Controlling the use of hash and URL certificate encoding" on page 1100.

Enabling the NSSD to generate hash and URL certificate encoding

To enable the NSSD to generate hash and URL certificate encoding, perform the following steps:

1. Export the certificates in CERTDER format from RACF.

   Use the RACDCERT EXPORT command with the CERTDER format option to create a data set that contains the binary DER encoding of a certificate on a key ring. If the HTTP server is running on the local system, copy the data set to the location specified by the CertificateURL parameter value. If the HTTP server is running on a remote system, transfer the data set to the appropriate location using a

utility such as FTP. For more details about the RACDCERT command, see z/OS Security Server RACF Command Language Reference.

Tip: Do not export the private key when you export the certificate from RACF.

2. Populate the HTTP server with exported certificates.
3. Identify the resources to the NSSD using the CertificateURL parameter.

   The CertificateURL parameter in the configuration file of the NSS server associates a certificate on the key ring of the NSS server with a URL that identifies an HTTP server and a file on that server that contains the binary DER encoding of the certificate. For more details about the CertificateURL parameter, see z/OS Communications Server: IP Configuration Reference.
4. Create the certificate bundles that you need by issuing the **certbundle** command.

   Use the **certbundle** command to create a file or data set that contains a certificate bundle. If the HTTP server is running on the local system, copy the file or data set to the location specified by the CertificateBundleURL parameter value. If the HTTP server is running on a remote system, transfer the file or data set to the appropriate location using a utility such as FTP. For more details about the **certbundle** command, see z/OS Communications Server: IP System Administrator's Commands. For more details about creating certificate bundles, see “Creating certificate bundles” on page 1100.
5. Populate the HTTP server with the certificate bundle files.
6. Identify the resources to the NSSD using the CertificateBundleURL parameter.

   The CertificateBundleURL parameter in the configuration file of the network security server associates a certificate on the key ring of the network security server with a URL that identifies an HTTP server and a file on that server that contains the certificate in a certificate bundle. For more details about the CertificateBundleURL parameter, see z/OS Communications Server: IP Configuration Reference.

Enabling the NSSD to process received hash and URL certificate encoding

To enable the NSSD to process received hash and URL certificate encoding, perform the following steps:

1. Ensure that HTTP traffic is not impeded by IP filter rules.

   Tip: If IP filtering is enabled on the system where the network security server is running, ensure that the correct filter rules are in place to allow communication with the HTTP servers that are identified on a CertificateURL or CertificateBundleURL, as well as any HTTP servers used by the remote security endpoint of the network security client. This communication typically uses the TCP protocol with an ephemeral source port and a destination port of 80.
2. Use the URLCacheInterval parameter on the IPSecDisciplineConfig statement in the NSSD configuration file to determine the maximum amount of time that URL data is cached before being re-fetched from an HTTP server. For more details about the URLCacheInterval parameter, see z/OS Communications Server: IP Configuration Reference.

Controlling the use of hash and URL certificate encoding

To control the use of hash and URL certificate encoding, configure IP security policies to accept hash and URL encoded certificates by setting the CertificateURLLookupPreference parameter on the KeyExchangePolicy and KeyExchangeAction statements in the IP security policy configuration file of network security clients. For more details about the CertificateURLLookupPreference parameter on the KeyExchangePolicy and KeyExchangeAction statements, see z/OS Communications Server: IP Configuration Reference.

Creating certificate bundles

Certificate bundles are used to store a group of related certificate information. A certificate bundle contains zero or more certificates and zero or more certificate revocation lists (CRLs). When an IKEv2 negotiation uses a digital signature authentication method, this certificate information can be exchanged using a certificate bundle. When information is exchanged using a certificate bundle, a URL that identifies the certificate bundle and a hash of the data in the certificate bundle is sent to the remote security

endpoint. The remote security endpoint then retrieves the certificate bundle from an HTTP server, and uses the bundle when it validates the digital signature.

A certificate bundle can hold in a single location all relevant information about an entire trust chain. The following types of information can be included in a certificate bundle:

- The certificate that was used to create a digital signature
- The certificates of certificate authorities in the trust chain
- Certificate revocation lists (CRLs)

Although consolidating this information in one place has advantages, consolidation might cause the remote security endpoint to retrieve unneeded information. Often the remote security endpoint already has knowledge of most of the certificates in the trust chain and is capable of retrieving CRL information using another method. In such cases, it might be more efficient to use individual certificates, rather than a certificate bundle.

You can use the **certbundle** command to create one or more files, each of which contains one certificate bundle. A certificate bundle options file is required as input to the **certbundle** command. The certificate bundle options file identifies how many certificate bundles are created, as well as the contents of each certificate bundle.

Guidelines:

- All certificate information in a certificate bundle file should be for the same trust chain.
- You should not include CRL information in a certificate bundle except when there is no other way for the remote security endpoint to retrieve the CRL information. Because certificate authorities periodically issue new CRLs, CRL information that is stored in a certificate bundle must be constantly updated to contain the most recent CRL information.

Rule: Do not put the certificate for the root certificate authority in a certificate bundle. An IKE implementation cannot accept a certificate for the root certificate authority from an untrusted source and, because certificate bundles are considered an untrusted source, any root certificates they contain are unusable. In addition, putting this certificate in the certificate bundle needlessly increases the size of the certificate bundle.

Steps for creating certificate bundles

You can use certificate bundles to consolidate all relevant information about an entire trust chain. The types of information that can be included in a certificate bundle are the certificate that was used to create a digital signature, the certificates of certificate authorities in the trust chain, and certificate revocation lists (CRLs).

Before you begin

Obtain from the certificate authority any certificate revocation lists (CRLs) that you want to put in a certificate bundle.

Procedure

Perform the following steps to create certificate bundles:

1. Store the CRLs that you are going to include in a certificate bundle in a file or data set.
2. Create a certificate bundle options file. See The z/OS UNIX certbundle command options file in z/OS Communications Server: IP System Administrator's Commands for more information.
3. For each certificate bundle that you are creating, define a CertBundleOptions statement:
   a) Use the KeyRing parameter to identify the key ring containing any certificates that you want to include.
   b) Use the CertificateChain parameter to specify the label of the certificate that is lowest in any complete trust chain that you want to include (excluding the root CA). The CertificateChain parameter generates a certificate bundle file that contains an optimal set of certificates.

c) Use the CertificateLabel parameter to specify the label of any individual certificates that you want to include. Use the CertificateLabel parameter only when you need to include fewer certificates than the entire chain.

d) Use the CRLFile parameter to identify the files that contain any CRLs that you want to include.

e) Use the BundleFile parameter to identify the name of the certificate bundle file that you are creating.

4. Provide read access to the key rings that are specified in the certificate bundle options file to the user ID under which the **certbundle** command is issued. See z/OS Security Server RACF Command Language Reference for details concerning access to key rings.
5. Issue the **certbundle** command, specifying the certificate bundle options file that you just created.

Controlling the NSS server

This topic describes starting and stopping the NSS server, modifying the configuration file, and displaying configuration file parameters.

Starting the NSS server

The NSS server can be started in the following ways:

- Using an MVS procedure from the MVS operator console. A sample start procedure is provided in SEZAINST(NSSD).
- From the z/OS shell, by starting OMVS and then issuing the **nssd** command.
- Using the COMMND*xx* member of PARMLIB. This allows the NSS server to be automatically started when the system is IPLed. For information about the use and configuration of the COMMND*xx* member of PARMLIB, see z/OS MVS Initialization and Tuning Reference.
- Using the AUTOLOG statement in the TCP/IP profile.

Tips:

- You should not start the NSS server using the AUTOLOG statement in a stack's profile. If the NSS server is listed in a stack's AUTOLOG statement, the server is cancelled if it is already running when that stack starts. This results in the NSS server losing any cached information that it has in place for NSS clients previously connected through all stacks, and could increase the overall recovery time when a TCP/IP stack recycles.
- If you start the NSS server from the z/OS shell and you stop the shell environment from scrolling, then when the **nssd** command needs to display data to the shell, the NSS server might stop and wait indefinitely for the shell to scroll and make output buffer space available for the data.
- When running from an MVS procedure, set the environment variables using the STDENV DD statement in the NSS server procedure.

Restriction: Only one instance of the NSS server can run on a z/OS image. If you attempt to start a second instance, the NSS server will fail.

Stopping the NSS server

Stop the NSS server from MVS by issuing the following command:

```
STOP procname
```

If the NSS server was started from a cataloged procedure, the *procname* value is the member name of that procedure. If the NSS server was started from the z/OS shell and the environment variable _BPX_JOBNAME was set, the *procname* value is the same as the _BPX_JOBNAME value. If the NSS server was started from the z/OS shell and _BPX_JOBNAME was not set, the *procname* value is based on the *userid*. If the *userid* is 8 characters long, the *procname* is the *userid*. If the *userid* is less than 8 characters

long, the *procname* is *useridX*, where *X* is the sequence number that is set by the system. To determine the sequence number, from the **ISPF LOG** window on TSO, issue the following command:

```
/d omvs,u=userid
```

This command displays the programs running under the specified user ID. For more information about _BPX_JOBNAME, see z/OS UNIX System Services Planning.

To stop the NSS server from the z/OS shell, issue the **kill** command (from a superuser ID) to the process ID (PID) that is associated with the NSS server. By default, the NSS server PID is recorded in /etc/nssd.pid. You can change the default location using the NSSD_PIDFILE environment variable.

Using the NSS server MODIFY command

The NSS server provides a modify command to take the following actions:

- Reread the configuration file.

  Use the MODIFY *procname*,REFRESH command to flush all cached URLs and reread the NSS server configuration file. Not all NSS server configuration parameters can be updated using this command. For information about which parameters can be dynamically changed, see the parameter descriptions for the NssConfig and IPSecDisciplineConfig statements in z/OS Communications Server: IP Configuration Reference.
- Display the configuration file parameters.

  Use the MODIFY *procname*,DISPLAY command to display configuration values currently in use by the NSS server.
- Display the contents of the URL cache.

  Use the MODIFY *procname*,DISPLAY,URLCACHE command to display the current contents of the URL data cache maintained by the NSS server.

For more information on the MODIFY command, see z/OS Communications Server: IP System Administrator's Commands.

NSS server failover considerations

NSS IPSec clients can use the NSS certificate service when negotiating phase 1 Security Associations. Network monitoring applications can use the NSS remote management service to display information about NSS IPSec clients. The NSS server should be treated as an application that requires high availability, an application that is able to recover quickly from an outage that impacts the ability of the NSS server to respond to IPSec clients.

Recovery configurations for the NSS server include:

- For recovery of NSS server workload by another NSS server within a sysplex, configure NSS IPSec clients to connect to the NSS server on a non-distributed dynamic VIPA. TCP/IP stacks configured as backup for the dynamic VIPA must have the necessary external security manager definitions and certificates to support the NSS IPSec clients, and an NSS server must be running on the z/OS system hosting the TCP/IP stack configured as backup.

  Guideline: Do not configure NSS IPSec clients to connect to a distributed DVIPA address on the NSS server. If a distributed DVIPA is used, the **ipsec** command and IPSec NMI can manage only NSS IPSec clients that have been distributed to the system on which the **ipsec** command is being run or to the system on which the IPSec NMI is invoked.
- Alternatively, you can configure an IKE daemon running as an NSS IPSec client to connect to a backup NSS server with the NetworkSecurityServerBackup parameter on the IkeConfig statement in the IKE daemon configuration file. When the IKE daemon is unable to connect to the primary NSS server, or when it loses its connection with the primary server, the IKE daemon attempts to connect to the server configured as backup. This recovery configuration can be used regardless of sysplex configurations. The backup server must be configured with all necessary external security manager definitions and

certificates to support the NSS IPSec clients. For additional details about the IkeConfig statement, see z/OS Communications Server: IP Configuration Reference.

NSS server capacity considerations

A single NSS server instance can support a maximum of 500 concurrent NSS client connections, in addition to 10 concurrent NMI client connections.

NSS server certificate revocation support

The NSS server supports the checking of certificate revocation lists (CRLs) when verifying a signature. The NSS server obtains the CRL of a certificate from an HTTP repository, determining the location of the CRL using the CRLDistributionPoints extension of the certificate. The NSS server searches each distribution point entry in the CRLDistributionPoints extension that contains a reason field pertaining to the all-reasons special value and that references the CRL by using an HTTP URL scheme. The search continues until a CRL that matches the certificate attributes is retrieved or until all distribution points containing an HTTP URL scheme are processed. If a certificate does not contain a CRLDistributionPoints extension or the CRLDistributionPoints extension does not contain at least one suitable distribution point that contains an HTTP URL scheme, then the NSS server is unable to retrieve the CRL.

The NSS server also supports the retrieval of certificate bundles, which can also contain a CRL. If a CRL cannot be retrieved using the CRLDistributionPoints extension of a certificate, the NSS server looks for a CRL in any certificate bundle that has hash and URL information provided by the network security client. The network security client obtains certificate bundle hash and URL information from certificate payloads sent by a remote security endpoint.

Managing network security services

Use the **ipsec** command to display information about NSS IPSec clients that are connected to the NSS server. You can also use this command to manage NSS IPSec clients that are enabled to use the NSS IPSec remote management service and that are currently connected to the NSS server.

Use the -x primary option on the **ipsec** command to display connection information about NSS IPSec clients connected to the NSS server.

```
ipsec -x display

CS V1R12 ipsec  NS Client Name: n/a  Mon Nov 27 12:40:02 2006
Primary:  NS Server       Function: Display            Format:   Detail
Source:   Server          Scope:    n/a                TotAvail: 1
SystemName: MVS052

ClientName:                      client4
ClientAPIVersion:                2
StackName:                       TCPCS4
SystemName:                      MVS052
ClientIPAddress:                 ::ffff:10.10.10.1
ClientPort:                      50003
ServerIPAddress:                 ::ffff:10.10.10.99
ServerPort:                      4159
UserID:                          USER1
RemoteManagementSelected:        Yes
RemoteManagementEnabled:         Yes
CertificateServicesSelected:     Yes
CertificateServicesEnabled:      Yes
ConnectState:                    connected
TimeConnected:                   2006/11/27 12:37:08
TimeOfLastMessageFromClient:     2006/11/27 12:37:08
*********************************************************************

1 entries selected
```

Use the **nssctl** command to display information about all of NSS clients that are connected to the NSS server.

```
nssctl -d
CS V1R12 nssctl  SystemName: MVS046    Mon Jun  9 17:05:16 2008
```

```
Function: Display            NSSClientName: n/a

ClientName:                      MVS046_TCPCS
ClientAPIVersion:                2
StackName:                       TCPCS
SystemName:                      MVS046
ClientIPAddress:                 ::ffff:9.42.105.149
ClientPort:                      50000
ServerIPAddress:                 ::ffff:9.42.105.149
ServerPort:                      4159
UserID:                          user1
ConnectState:                    connected
TimeConnected:                   2008/06/09 12:22:32
TimeOfLastMessageFromClient:     2008/06/09 12:22:48
Discipline:                      IPSec
  CertificateServiceSelected:    Yes
  CertificateServiceEnabled:     Yes
  RemoteManagementSelected:      Yes
  RemoteManagementEnabled:       Yes
*************************************************************
ClientName:                      XMLAllClient1
ClientAPIVersion:                3
StackName:                       Any
SystemName:                      dpsys01
ClientIPAddress:                 ::ffff:9.42.105.149
ClientPort:                      1026
ServerIPAddress:                 ::ffff:9.42.105.149
ServerPort:                      4159
UserID:                          USER1
ConnectState:                    connected
TimeConnected:                   2008/06/09 17:05:11
TimeOfLastMessageFromClient:     2008/06/09 17:05:11
Discipline:                      XMLAppliance
  CertificateServiceSelected:    Yes
  CertificateServiceEnabled:     Yes
  PrivateKeyServiceSelected:     Yes
  PrivateKeyServiceEnabled:      Yes
  SAFAccessServiceSelected:      Yes
  SAFAccessServiceEnabled:       Yes
*************************************************************

2 entries selected
```

Use the -z option on the **ipsec** command to specify the name of an NSS client rather than a name of a local TCP/IP stack. When the -z option is specified, the **ipsec** command obtains information about the NSS client from the NSS server. The -z option is valid only on the system that is running the NSS server. The NSS client that is identified by the -z option must be connected to the NSS server. The NSS client must also be enabled to use the NSS remote management service. The following example uses the -z option to display phase 2 Security Association information about the NSS client client4, where the name client4 was obtained from the previous **ipsec -x display** command.

```
ipsec -y display -z client4

CS V1R12 ipsec  NS Client Name: client4  Mon Nov 27 12:44:35 2006
Primary:  Dynamic tunnel  Function: Display           Format:   Detail
Source:   Stack           Scope:    Current           TotAvail: 1

TunnelID:                     Y2
Generation:                   1
IKEVersion:                   1.0
ParentIKETunnelID:            K1
VpnActionName:                Dvpn
LocalDynVpnRule:              mvs052_192
State:                        Active
HowToEncap:                   Tunnel
LocalEndPoint:                10.10.10.1
RemoteEndPoint:               10.10.10.2
LocalAddressBase:             10.10.10.1
LocalAddressPrefix:           n/a
LocalAddressRange:            n/a
RemoteAddressBase:            10.10.10.2
RemoteAddressPrefix:          n/a
RemoteAddressRange:           n/a
HowToAuth:                    AH
 AuthAlgorithm:               Hmac_Sha
 AuthInboundSpi:              2401615039
 AuthOutboundSpi:             1971620597
HowToEncrypt:                 3DES
```

```
 EncryptInboundSpi:           4088723240
 EncryptOutboundSpi:          445063417
Protocol:                     ALL(0)
LocalPort:                    0
LocalPortRange:               n/a
RemotePort:                   0
RemotePortRange:              n/a
Type:                         n/a
TypeRange:                    n/a
Code:                         n/a
CodeRange:                    n/a
OutboundPackets:              0
OutboundBytes:                0
InboundPackets:               0
InboundBytes:                 0
Lifesize:                     0K
LifesizeRefresh:              0K
CurrentByteCount:             0b
LifetimeRefresh:              2006/11/27 14:09:19
LifetimeExpires:              2006/11/27 14:44:19
CurrentTime:                  2006/11/27 12:44:35
VPNLifeExpires:               2007/03/07 12:44:19
NAT Traversal Topology:
  UdpEncapMode:               No
  LclNATDetected:             No
  RmtNATDetected:             No
  RmtNAPTDetected:            No
  RmtIsGw:                    n/a
  RmtIsZOS:                   n/a
  zOSCanInitP2SA:             n/a
  RmtUdpEncapPort:            n/a
  SrcNATOARcvd:               n/a
  DstNATOARcvd:               n/a
  PassthroughDF:              No
  PassthroughDSCP:            No
*******************************************************************************

1 entries selected
```

For details about the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Chapter 19. Defensive filtering

An external security information and event manager, by analyzing and correlating messages from multiple sources and systems in the network, can take action to block attacks by installing defensive filters in your TCP/IP stack. A defensive filter is an IP filter rule to discard packets, separate from IP security filters, and is typically installed for a short duration (for example, 30 minutes) to block a specific attack or a pattern of attacks. If traffic being blocked by a defensive filter should be blocked on a long-term basis, update your configured IP security policy to add an IP security deny rule.

A defensive filter uses a combination of the following characteristics to target traffic to be discarded:

- IP source or destination address
- IP protocol
- Source or destination port
- ICMP type or code
- Direction of flow
- Type of traffic: Routed or local

For example, a defensive filter might be installed to block all TCP traffic from IP address 10.1.1.1 that is destined for the Telnet server. The characteristics of this filter are the following characteristics:

- IP source address is 10.1.1.1.
- IP protocol is TCP.
- Destination port is 23.
- Direction of flow is inbound.
- Traffic is local.

Defensive filters are given higher priority than IP security filters. That is, IP filter processing first checks any installed defensive filters for a match against a packet, before checking the IP security filters. When a defensive filter is added to a TCP/IP stack, it is placed at the top of the filter search order.

Figure 138 on page 1108 provides an overview of defensive filtering.

© Copyright IBM Corp. 2000, 2022

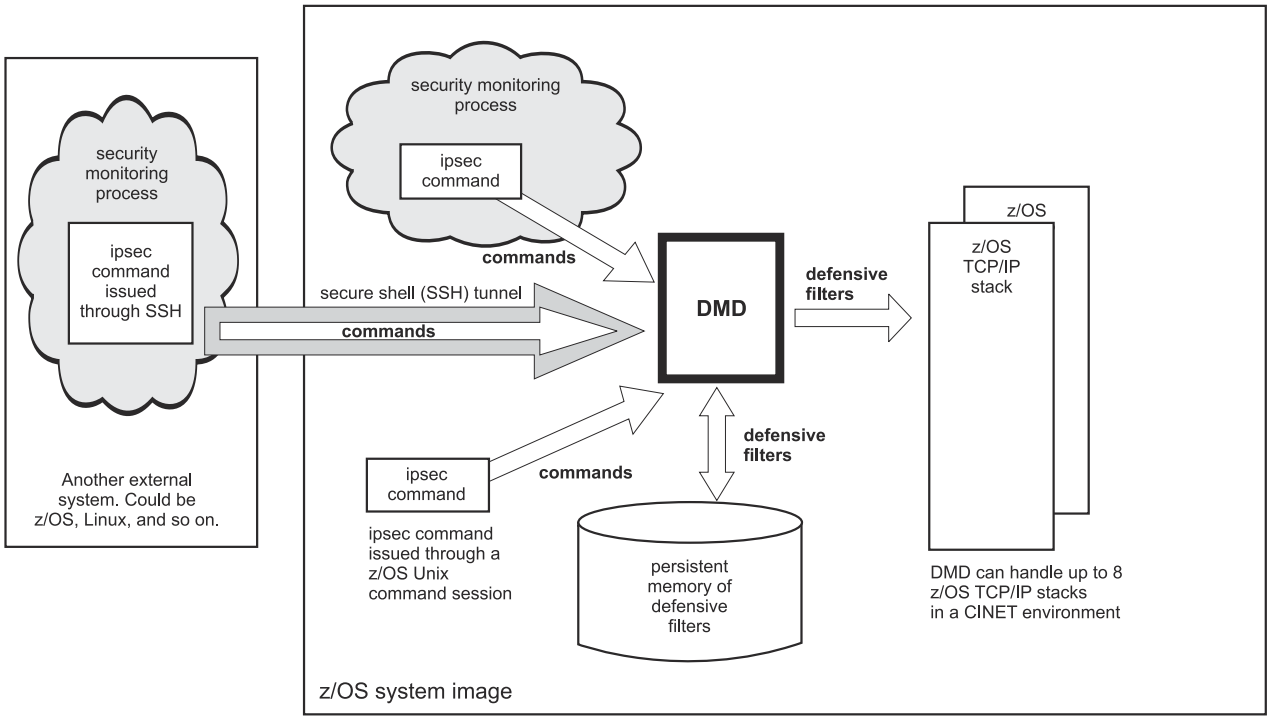


Figure 138. Defensive filtering overview

Defensive filters are added and managed using the z/OS UNIX **ipsec** command with the -F primary option.

- Defensive filters are typically added as an automated action resulting from an external security information and event manager's analysis. The manager issues the set of **ipsec** commands that install the required defensive filters.
- You can also add a defensive filter by manually issuing the **ipsec** command.
- After a defensive filter is created, you can use the **ipsec** command to update some attributes of the filter, such as its lifetime, and also to display and delete defensive filters.

For more information about the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Requirements:

- You must enable the IP security function for defensive filters to be installed in a stack. If you do not have the IP security function enabled, see "Enabling the IP security function" on page 1115.
- The Defense Manager daemon (DMD) plays an integral role in managing defensive filters, and must be active for defensive filters to be added, updated, or deleted. One instance of the DMD manages all eligible stacks on a z/OS image. An eligible stack is one that is enabled for IP security and that is included in the DMD configuration file with a mode of Active or Simulate. For information about configuring the DMD, see "Steps for configuring the DMD" on page 1116. You can refresh most of the DMD configuration parameters so that options can be changed without recycling the DMD.

Guideline: The DMD can support a maximum of 10 concurrent **ipsec** command connections.

Restriction: Remote management of defensive filters using a network security services (NSS) server is not supported. Management of defensive filters is provided only through the local **ipsec** command.

Global and stack-specific defensive filters

A defensive filter has either a global or stack-specific scope.

- A global defensive filter, identified by the -G option on the **ipsec** command, is installed in all eligible stacks on the z/OS image.

- A stack-specific defensive filter is installed in the stack specified by the -p *stackname* option on the **ipsec** command, if the stack is eligible.

When a defensive filter is created, the DMD installs the defensive filter in the eligible stacks and maintains a copy of the filter. The DMD maintains a persistent record of all active defensive filters, global and stack-specific, and the stacks into which those filters are installed.

This persistence enables the DMD to maintain the correct set of installed defensive filters across startup and shutdown activities for TCP/IP stacks, as well as across startup and shutdown of the DMD itself. If the DMD is already running when an eligible TCP/IP stack starts, the DMD installs all applicable defensive filters, global and stack-specific, to that stack. If, when the DMD starts up, it detects that eligible stacks are already running, it ensures that they have or receive the correct set of defensive filters.

After a global defensive filter is created and installed in one or more stacks, you can update it as a global filter, resulting in all copies of the filter being updated. You can also update it in an individual TCP/IP stack, resulting in only that stack's copy of the filter being updated. Similarly, you can delete a global defensive filter globally or only from an individual TCP/IP stack.

Defensive filter names

A defensive filter receives its name from the -N *DefensiveFilterName* option when the filter is created. All copies of a global defensive filter have the same name. Global filters and stack-specific filters share the same filter namespace, so a filter name cannot be used for both a global filter and a stack-specific filter. The creation of a stack-specific filter fails if the filter name conflicts with a global filter of the same name, or if there is already a stack-specific filter of the same name in the target stack. The creation of a global filter fails if the filter name conflicts with a global filter of the same name, or if there is already a stack-specific filter of the same name in any stack. This check includes filter names that are in the DMD's persistent memory, even if the corresponding stack is not active.

Tip: If you are manually creating defensive filters, avoid conflicts between global and stack-specific filter names by choosing a distinct naming convention for each, such as starting all global filter names with a G.

Defensive filter modes

Each defensive filter has a mode setting of block or simulate. The defensive filter's mode is set when the filter is created or updated by the **ipsec** command.

By default, defensive filters are in block mode, causing traffic to be discarded. A defensive filter in simulate mode simulates a block and lets you monitor the impact of enabling defensive filters without discarding traffic.

When a packet matches a defensive filter and the mode is simulate, a message is logged indicating that the packet would have been discarded, but the packet is not discarded and IP filtering continues. The packet can subsequently match a defensive filter that is in block mode and be discarded, but the packet will not match another simulation filter.

The DMD configuration file also provides the mode settings Active, Simulate, or Inactive on the DmStackConfig statement.

- Active enables defensive filtering and honors the mode setting of the individual filters.
- Simulate enables defensive filtering and overrides the mode setting of the individual filters; simulate mode is used for all defensive filters installed in the stack.
- Inactive disables defensive filtering.

Table 56 on page 1110 summarizes the interaction between the mode setting on the DmStackConfig statement and the mode setting in individual filters set by the **ipsec** command.

Table 56. Interaction between the mode setting on the DmStackConfig statement and the mode setting in individual filters				
		Mode setting on the DmStackConfig statement		
		Active	**Simulate**	**Inactive**
Individual filter mode set by the ipsec command	**Block**	Block the packet	Simulate blocking the packet	No defensive filters
	Simulate	Simulate blocking the packet	Simulate blocking the packet	No defensive filters

Tips:

- You might want to specify Mode Simulate on the DmStackConfig statement when you are first implementing defensive filtering. All defensive filters in the TCP/IP stack will be treated as if the mode was simulate. When a packet matches a defensive filter, syslog message EZD1722I is generated and IP filtering continues. Defensive filters added to this stack retain the mode setting with which they were added, block or simulate. In most cases, you should use the default mode, block, on the individual filter.
- After completing defensive filter testing in simulate mode, specify Mode Active on the DmStackConfig statement. If there are defensive filters installed in the stack when the mode is changed from simulate to active, the mode on the individual defensive filters is used.
- If defensive filtering is active (DmStackConfig statement with Mode Active) and you want to implement and test additional automation, you can revert to an overall mode of simulate for the whole stack. However, you might want only defensive filters added by the new automation to have a mode of simulate. The automation action can add individual defensive filters with a mode of simulate using the mode keyword on the **ipsec -F add** command. After testing, you can update the automation action to add defensive filters with a mode of block using the mode keyword on the **ipsec -F add** command.

For more information about the DmStackConfig statement, see z/OS Communications Server: IP Configuration Reference. For more information about adding or updating a defensive filter with the -F option of the **ipsec** command, see z/OS Communications Server: IP System Administrator's Commands.

Allowing administrative access

Defensive filters are checked before IP security filters. To ensure that an administrator is not blocked by a defensive filter, you can exclude the administrator's IP address from defensive filter processing by specifying the administrator's address on the Exclude parameter of the DmStackConfig statement in the DMD configuration file. For more information about the DmStackConfig statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

Filter-match logging

When a packet matches a defensive filter during IP filter processing, a message can be logged indicating that the packet was discarded based on this filter. When a defensive filter is added, filter-match logging can be enabled or disabled for the filter; it is enabled by default. The filter-match logging setting can be updated with the **ipsec** command for an existing defensive filter.

When a defensive filter is simulating a block, filter match logging is always performed to indicate that a packet would have been discarded based on the defensive filter.

You can limit the number of filter-match messages generated for a defensive filter by specifying a log limit for the defensive filter in one of the following ways:

- Use the loglimit keyword on the **ipsec** command when you add a defensive filter.
- Use the loglimit keyword on the **ipsec** command to update a defensive filter.
- Set a default value for one or more TCP/IP stacks using the DefaultLogLimit parameter on the DmStackConfig statement in the DMD configuration file. The DefaultLogLimit value is used when a filter

is added without specifying a loglimit value. For more information about the DmStackConfig statement and its parameters, see z/OS Communications Server: IP Configuration Reference.

The log limit limits the average rate of filter-match messages generated in a 5-minute interval for a defensive filter. For example, a loglimit value of 100 limits the average rate of filter-match messages to 100 messages per 5-minute interval. A burst of up to 100 messages is allowed while maintaining the long-term average of 100 messages per 5-minute interval.

A count is kept of suppressed messages. At the end of the 5-minute interval, if there are suppressed filter-match messages, EZD0837I or EZD0838I is generated to report the number of suppressed messages.

TRMD

The Traffic Regulation Manager daemon (TRMD) is responsible for logging defensive filter events that are detected by the stack. These events include filter-match logging and the creation, deletion, and updating of defensive filters.

TRMD and syslogd provide the logging service for defensive filtering. In a Common INET environment, you must configure one instance of TRMD for each stack on a z/OS system. For information about configuring TRMD, see "TRMD" on page 879.

Disabling defensive filters for a single stack

To disable defensive filtering for a TCP/IP stack, while you continue to support defensive filtering for other stacks on the system, take the following actions:

1. Update the DMD configuration file and change the mode of the stack to Inactive on the DmStackConfig statement.
2. Issue the MODIFY REFRESH command for the DMD.

Results:

- All defensive filters are removed from the stack.
- The DMD's persistent memory of defensive filters for the stack is cleared.
- Additional defensive filters cannot be added to this stack.

Tips:

- If you cannot update your DMD configuration file, you can issue the MODIFY FORCE_INACTIVE command for the DMD to disable defensive filtering for the stack. However, a later MODIFY REFRESH command will use the DMD configuration file, so if you want defensive filtering to remain disabled, you should update the DMD configuration file as soon as possible.
- Removing the DmStackConfig statement from the DMD configuration file does not delete existing defensive filters from the stack. If you remove the DmStackConfig statement, the defensive filters remain in the stack until they expire. To remove the defensive filters from the stack immediately, add the DmStackConfig statement back to the DMD configuration file and specify mode Inactive, or issue the MODIFY FORCE_INACTIVE command for the stack.

Relationship between intrusion detection services and defensive filters

Communication Server's intrusion detection services (IDS) support enables you to detect scans of your TCP/IP stack and possible attacks. It also provides traffic regulation for TCP connections and UDP sockets. One action that can be taken when a scan or attack is detected, or traffic regulation is enforced, is to generate a message to report the event.

An external security information and event manager that is configured to receive messages from the TCP/IP stack's IDS function can analyze the messages and correlate the information with other information that it has received. Communication Server's IDS messages can be one of a number of inputs

that an external security information and event manager uses to make the decision to add a defensive filter to the stack. If the external security information and event manager detects an attack, it can add defensive filters to the stack to block the attack. Defensive filter support can be enabled without enabling Communication Server's IDS support.

For more information about IDS support, see Chapter 16, “Intrusion detection services,” on page 851.

Comparison of IP security filters and defensive filters

Table 57 on page 1112 compares IP security filters and defensive filters.

Table 57. Comparison of IP security filters and defensive filters

Topic	IP security filters (policy)	IP security filters (default)	Defensive filters
Configuring	Configured in a Policy Agent flat file.	Configured in the TCP/IP profile.	Not configured. The **ipsec** command is used to create defensive filters, either automatically or manually.
Installing in the TCP/IP stack	Installed by the Policy Agent.	Installed by TCP/IP profile processing.	Installed by the Defense Manager daemon (DMD).
Filter search order	The order in the configuration file.	The order in the configuration file.	Defensive filters are searched before IP security filters. When a defensive filter is created, it is installed at the top of the search order.
Displaying a filter	Use **pasearch** and **ipsec -f display**. The **ipsec -f display -c current** command displays all installed filters, both defensive filters and IP security filters.	Use **ipsec -f display -c profile**.	Use **ipsec -F display**.

Table 57. Comparison of IP security filters and defensive filters (continued)

Topic	IP security filters (policy)	IP security filters (default)	Defensive filters
Filter display order	The order in the configuration file. The **pasearch** command displays IP security filters as complex filter rules, not split filters as they are in the stack. The **ipsec -f display** command displays IP security filters as split filters, like they are in the stack. IPv4 IP security filters are shown first, followed by IPv6 IP security filters.	The order in the configuration file. The **ipsec -f display** command displays IP security filters as split filters, like they are in the stack. A single profile filter in the configuration file is split into an inbound and outbound filter in the stack. IPv4 IP security filters are shown first, followed by IPv6 IP security filters.	The **ipsec -F display** command displays defensive filters from the stack in four groups: • IPv4 inbound filters • IPv4 outbound filters • IPv6 inbound filters • IPv6 outbound filters Within each group, the filters are displayed from most recently installed to least recently installed. The **ipsec -F display -G** command displays global defensive filters from the DMD. The global filters are displayed from most recently installed to least recently installed.
Deleting a filter	Remove the filter rule from the configuration file. When Policy Agent detects the configuration file change, the filter rule is removed from the stack. Policy Agent detects the change in one of the following ways: • If Policy Agent was started with the `-i` startup option, an immediate refresh picks up the change. • You issue a MODIFY REFRESH command. • You issue a MODIFY UPDATE command. • Policy Agent checks for configuration changes using an update interval defined in the policy configuration file.	Use a VARY TCPIP,,OBEYFILE command with a data set that contains a new IPSEC statement with the filter rule removed.	Use **ipsec -F delete**. Defensive filters are also deleted when their lifetime expires.

Table 57. Comparison of IP security filters and defensive filters (continued)			
Topic	**IP security filters (policy)**	**IP security filters (default)**	**Defensive filters**
Updating a filter	Update the filter rule in the configuration file. When the Policy Agent detects the configuration file change, the filter rule is updated in the stack.	Use a VARY TCPIP,,OBEYFILE command with a data set that contains a new IPSEC statement with the filter rule updated.	Use **ipsec -F update**. A defensive filter's lifetime, mode, and logging values can be updated.
Specifying time conditions	Specify time conditions in the policy. The Policy Agent installs an IP security filter when it becomes active, and deletes the filter when it becomes inactive due to time.	Not supported.	Not supported. Defensive filters have a lifetime that is minutes in length. A defensive filter is deleted when its lifetime expires.
Simulation mode	Not supported.	Not supported.	Controlled by the DMD configuration file and the **ipsec -F add** and **ipsec -F update** commands.
Global filters	IP security filters defined in a CommonIPSecConfig file are added to all eligible stacks.	Not supported.	Defensive filters added with the -G option of the **ipsec** command are added to all eligible stacks on the z/OS system.
Filter-match logging	• Controlled by settings in the policy flat file. • Message generated for each packet that matches the filter, if logging enabled.	• Controlled by settings in the TCP/IP profile. • Message generated for each packet that matches the filter, if logging enabled.	• Set when the filter is added or updated with the **ipsec** command. • Messages can be limited by using the loglimit parameter on the **ipsec -F add** or **ipsec -F update** commands, or by using the DefaultLogLimit parameter in the DMD configuration file.

The DMD run-time environment

The DMD is a z/OS UNIX application; it requires the z/OS UNIX file system. You can start the DMD from an MVS started procedure, from the z/OS UNIX shell, with the AUTOLOG statement in the TCP/IP profile, or by using the COMMND*xx* member of parmlib. The DMD must be started by a RACF-authorized user ID, and it must be in an APF-authorized library. For more information about how to start the DMD, see "Starting the DMD" on page 1118.

The DMD uses the MVS operator's console, syslogd, CTRACE, and STDOUT for its logging and tracing. The MVS operator's console and STDOUT are used for major events such as initialization, termination, and error conditions. Syslogd is used for logging events related to the processing of defensive filter requests. CTRACE is used for detailed tracing and debugging.

The DMD uses a standard message catalog. The message catalog must be in the UNIX file system. The directory location for the message catalog path is set by the environment variables NLSPATH and LANG.

The DMD and Language Environment run-time options

When you start the DMD from a started or cataloged procedure, you typically start it directly from the SEZALOAD data set using PGM=EZADMD. However, you can also start the DMD using BPXBATCH.

When you start the DMD using PGM=EZADMD, the STDENV DD card, if used, is passed directly to the DMD. Language Environment does not have access to the STDENV environment variables. As a result, any Language Environment run-time options set in the STDENV DD data set using the _CEE_RUNOPTS environment variable are ignored. In this case, Language Environment run-time options must be passed on the PARM parameter, and the options must be specified before any DMD options. However, the PARM parameter allows a maximum of 100 characters. If the Language Environment run-time options plus the DMD parameters that you want to specify exceed 100 characters, consider using BPXBATCH to start the DMD. When you use PGM=BPXBATCH to start the DMD, you can include the Language Environment variable _CEE_RUNOPTS on the STDENV DD card to specify run-time options in excess of 100 characters long.

For more information about how to start the DMD, see “Starting the DMD” on page 1118.

Enabling defensive filtering

To enable defensive filtering, first take both of the following actions:

- Enable the IP security function.

  See “Enabling the IP security function” on page 1115.
- Configure the Defense Manager daemon (DMD).

  See “Steps for configuring the DMD” on page 1116.

After configuring the DMD and enabling IP security, start the DMD. For details, see “Starting the DMD” on page 1118.

Enabling the IP security function

The IP security function must be enabled for defensive filters to be installed in a stack. If you do not have the IP security function enabled and want to use defensive filters, perform the following steps.

1. Specify the IPSECURITY parameter on the IPCONFIG statement in the TCP/IP profile.
2. If you have IPv6 traffic in your network and want to use defensive filters for IPv6, specify the IPSECURITY parameter on the IPCONFIG6 statement in the TCP/IP profile.
3. Configure a default IP security filter policy in the TCP/IP profile using the IPSEC statement in the TCP/IP profile.

   Tip: To permit all IPv4 traffic, you can configure a single rule for the IPSEC statement that allows all traffic. The following example allows all IPv4 traffic, without logging filter matches.

   ```
   IPSEC
   ; Rule   SourceIp      DestIp    Logging   Prot       SrcPort    DestPort   Routing        Secclass
   ;
   ; Permit all local and routed IPv4 traffic, no logging.
     IPSECR *             *         NOLOG     PROTO *                          ROUTING EITHER
   ENDIPSEC
   ```

   For more information about the IPSEC statement, see z/OS Communications Server: IP Configuration Reference.
4. Optionally, configure a more comprehensive IP security policy in a flat file for the Policy Agent to install and manage.

   You can configure IPSec encryption and authentication only in a Policy Agent flat file.

Tip: You can use the IBM Configuration Assistant for z/OS Communications Server to create an IP security policy flat file.

For more information about configuring an IP security policy, see Chapter 14, "Policy-based networking," on page 787 and Chapter 17, "IP security," on page 883.

Attention: You must configure an IP security policy if you enable the IP security function. If you specify IPSECURITY in your TCP/IP profile and do not configure an IP security policy, all inbound and outbound traffic will be discarded.

Steps for configuring the DMD

The Defense Manager daemon (DMD) plays an integral role in managing defensive filters, and must be active for defensive filters to be added, updated, or deleted.

Procedure

Perform the following steps to configure the DMD:

1. Authorize the DMD to the external security manager.

   See "Steps for authorizing resources for the DMD and the ipsec command" on page 1118.
2. Create the directories that the DMD needs.
   a) Create the directory /var/dm for use by the DMD. The DMD user ID must have permission to create, delete, read, and write files to this directory.
   b) If you set the PID file location with the DMD_PIDFILE environment variable, ensure that the path portion of the file name exists and that the DMD user ID has permission to create and write files to that directory. If you use the default PID file location, /var/dm/dmd.pid, you have already created the directory and given the DMD user ID the appropriate access in the previous step.
   c) Create the directory that will hold the persistent defensive filters for each stack, as well as the global defensive filters.

      The DMD configuration file parameter DefensiveFilterDirectory points to this directory. The default value is /var/dm/filters. Ensure that the DMD user ID is authorized to create, delete, read from, and write to files in this directory. The directory should have sufficient space to support at least 1 MB of data for each TCP/IP stack, plus another 1 MB for the global filter definitions. For more information about the DefensiveFilterDirectory parameter in the DMD configuration file, see z/OS Communications Server: IP Configuration Reference.
3. Create the DMD configuration file.

   Take one of the following actions:

   - Use the IBM Configuration Assistant for z/OS Communications Server.

     The IBM Configuration Assistant for z/OS Communications Server, an optional GUI-based tool, provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the DMD configuration file.

     The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

     Through a series of wizards and online help panels, you can use the GUI to create DMD configuration files for any number of z/OS images with any number of TCP/IP stacks per image.
   - Configure the file manually.

     A sample configuration file is in /usr/lpp/tcpip/samples/dmd.conf.

     For a description of the DMD configuration file, see z/OS Communications Server: IP Configuration Reference.

If the DMD was defined to the external security manager with a nonzero UID, ensure that the DMD has permission to read the configuration file. The DMD user ID must have both read access to the configuration file and execute access to the directory containing the configuration file.

Tip: You can create the configuration file in the /var/dm directory and use the DMD_FILE environment variable to specify the configuration file. You set up the /var/dm directory in step 2 to allow DMD to create, delete, read, and write files to this directory.

The following search order is used by the DMD to locate the configuration data set or file:

a. If the environment variable DMD_FILE is defined, the DMD uses the value as the name of an MVS data set or z/OS UNIX file to access the configuration data.

b. /etc/security/dmd.conf

You can specify statements in the configuration file using a variety of EBCDIC code pages. Use the DMD_CODEPAGE environment variable to specify the code page that you want to use. The default code page is IBM-1047.

4. Optionally, set the _BPX_JOBNAME environment variable.

   When you start the DMD from the z/OS UNIX shell, set the environment variable _BPX_JOBNAME. This enables a specific job name to be used with the STOP or MODIFY console commands. For information about _BPX_JOBNAME, see z/OS UNIX System Services Planning.

5. Configure and start syslogd.

   The DMD uses the local4 facility when writing messages to syslogd. For performance purposes, syslogd should use z/OS File System as its underlying file system. For more information about syslogd, see "Configuring the syslog daemon" on page 213.

6. Optionally, update the DMD environment variables.

   The DMD uses the following environment variables. You can modify them for your installation.

   DMD_CODEPAGE
   Use the DMD_CODEPAGE variable to specify the EBCDIC code page to be used when reading the configuration file. For details about the supported code pages, see z/OS Communications Server: IP Configuration Reference.

   DMD_CTRACE_MEMBER
   Used by the DMD to locate a parmlib member for DMD CTRACE customization. For more information about the TCP/IP services component trace for the DMD, see z/OS Communications Server: IP Diagnosis Guide.

   DMD_FILE
   Used by the DMD in the search order for the DMD configuration file. For details about the search order used for locating this configuration file, see step "3" on page 1116.

   DMD_PIDFILE
   Used by the DMD in the search order for the file that should contain the DMD process ID (PID). The search order for the DMD PID file is as follows:

   a. DMD_PIDFILE environment variable

   b. /var/dm/dmd.pid

7. If you are starting the DMD as a started procedure, update the DMD cataloged procedure.

   Create the cataloged procedure by copying the sample in SEZAINST(DMD) to your system. Specify the DMD parameters and change the data set names to suit your local configuration. A copy of the DMD cataloged procedure can also be found in z/OS Communications Server: IP Configuration Reference.

   If the DMD was defined to the external security manager with a nonzero UID and the cataloged procedure specifies an HFS file containing environment variables, ensure that the DMD has permission to read the HFS file.

Results

You know you are done when you can start the DMD. For details, see "Starting the DMD" on page 1118.

Steps for authorizing resources for the DMD and the ipsec command

You need to define and authorize the DMD user ID, permit the DMD user ID to SYS1.PARMLIB, and define the SERVAUTH profiles necessary to be able to add, update, delete, and display defensive filters

Before you begin

RACF is used as the external security manager in the following examples. However, you can use any SAF-compliant security product. RACF commands shown in these examples are also provided in the EZARACF member of the SEZAINST data set. In these examples, it is assumed that the DMD is running under the user ID DMD.

Procedure

Perform the following steps to authorize access to the appropriate resources:

1. Define and authorize the DMD user ID.

   The DMD is a z/OS UNIX application that you can start from the z/OS UNIX shell or from an MVS started procedure. Before starting the DMD, you must define the DMD user ID to the external security manager. If you start the DMD from an MVS started procedure, the DMD user ID must also be authorized to the STARTED class. In the following example, the DMD user ID is defined with UID 0:

   ```
   ADDUSER  DMD     DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(0)  HOME('/'))
   RDEFINE  STARTED  DMD.*            STDATA(USER(DMD))
   SETROPTS RACLIST(STARTED) REFRESH
   SETROPTS GENERIC(STARTED) REFRESH
   ```

   You can define the DMD with a nonzero UID. For additional steps that you must take when the DMD UID is nonzero, see “Steps for configuring the DMD” on page 1116.

2. Permit the DMD user ID to SYS1.PARMLIB.

   The DMD uses the TCP/IP component trace (CTRACE) to perform service-level tracing. The default DMD component trace parmlib member is stored in SYS1.PARMLIB. The DMD user ID must be permitted to access SYS1.PARMLIB.

   Issue the following command:

   ```
   PERMIT   SYS1.PARMLIB  ID(DMD) ACCESS(READ)
   ```

3. Define SERVAUTH profiles to control the users that are allowed to manage defensive filters.

   For information about defining the SERVAUTH profiles needed for a user to be able to add, update, delete, and display defensive filters, see “Step 4: Authorizing the ipsec command to the external security manager” on page 1363. Additional information about command security and the SERVAUTH profile is available in z/OS Communications Server: IP System Administrator's Commands.

Starting the DMD

You can start DMD in any of the following ways:

- By using an MVS procedure from the MVS operator console. A sample start procedure is provided in SEZAINST(DMD).
- By issuing the **dmd** command from the z/OS UNIX shell.
- By using the COMMND*xx* member of parmlib. This member enables the DMD to be automatically started after an IPL of the system. For information about configuring and using the COMMNDxx member of parmlib, see z/OS MVS Initialization and Tuning Reference.
- By using the AUTOLOG statement in the TCP/IP profile. For information about the AUTOLOG statement, see z/OS Communications Server: IP Configuration Reference.

Tips:

- Do not start the DMD using the AUTOLOG statement in a stack's profile if you are running in a CINET environment with more than one stack configured. If a stack is configured using AUTOLOG to start and stop the DMD each time the stack starts and stops, it is difficult to maintain a stable running instance of the DMD in a multi-stack environment. In a multi-stack environment, you should use another method to automate starting the DMD when the system is started, such as using the COMMND*xx* member of parmlib.
- When you start the DMD from an MVS procedure, set the environment variables using the STDENV DD statement in the DMD procedure.
- If you start the DMD from the z/OS shell and you stop the shell environment from scrolling, when the daemon needs to display data to the shell, it might stop and wait indefinitely for the shell to scroll and make output buffer space available for the data.

Restriction: Only one instance of the DMD can run on a z/OS image. If you attempt to start a second instance, the second DMD will fail.

Stopping the DMD

Stop the DMD in one of the following ways:

- From MVS, stop the DMD by issuing the following command:

  ```
  STOP procname
  ```

  where *procname* is one of the following values:

  – If the DMD was started from a cataloged procedure, the *procname* value is the member name of that procedure.

  – If the DMD was started from the z/OS UNIX shell and the environment variable _BPX_JOBNAME was set, the *procname* value is the same as the _BPX_JOBNAME value.

  – If the DMD was started from the z/OS UNIX shell and _BPX_JOBNAME was not set, the *procname* value is based on the *userid*. If the *userid* is 8 characters long, the *procname* is the *userid*. If the *userid* is less than 8 characters long, the *procname* is *useridX*, where *X* is the sequence number that is set by the system. To determine the sequence number, from the **ISPF LOG** window on TSO, issue the following command:

  ```
  /d omvs,u=userid
  ```

  This command displays the programs that are running under the specified user ID. For more information about _BPX_JOBNAME, see z/OS UNIX System Services Planning.
- From the z/OS shell, issue the **kill** command from a superuser ID for the process ID (PID) that is associated with the DMD. By default, the DMD PID is recorded in /var/dm/dmd.pid. You can change the default location using the DMD_PIDFILE environment variable.

Using the DMD MODIFY command

The DMD provides a MODIFY command to take the following actions:

- Reread the configuration file.

  Use the MODIFY *procname*,REFRESH command to reread the DMD configuration file. You can use this command to update some DMD configuration file parameters. For information about the DMD configuration file parameters that can be dynamically changed, see z/OS Communications Server: IP Configuration Reference.
- Display the configuration file parameters.

  Use the MODIFY *procname*,DISPLAY command to display configuration values currently in use by the DMD.

- Disable defensive filtering.

  Use the MODIFY *procname*,FORCE_INACTIVE,*stackname* command to disable defensive filtering for stack *stackname*. All defensive filters for the stack are removed from the DMD's persistent memory and from the stack. No additional defensive filters are added to the stack while the stack's mode is inactive. The change to the stack's mode persists until the next successful MODIFY *procname*,REFRESH command.

For more information about the MODIFY command for DMD, see z/OS Communications Server: IP System Administrator's Commands.

Chapter 20. Application Transparent Transport Layer Security data protection

The Transport Layer Security (TLS) protocol evolved from the Secure Sockets Layer (SSL) protocol and was originally defined in RFC 2246. The protocol enables client and server applications to communicate in a way that is designed to prevent eavesdropping, tampering, and message forgery. To implement TLS protocols, typically, applications must be modified to incorporate a TLS toolkit. Modifying applications requires significant development overhead, ongoing maintenance for each application, and application-specific knowledge of the parameters needed to implement TLS for that application.

Application Transparent Transport Layer Security (AT-TLS) consolidates TLS implementation in one location, reducing or eliminating application development overhead, maintenance, and parameter specification. AT-TLS is based on z/OS System SSL, and transparently implements these protocols in the TCP layer of the stack. As shown in Figure 139 on page 1121, most applications do not need any awareness of the security negotiations and encryption done by TCP/IP on its behalf. However, you might want some applications to be aware of AT-TLS or have control over the security functions being performed by TCP/IP. For example, if the application is a server requesting client authentication, you might want the application to get the partner certificate or the user ID associated with the partner certificate, or the application might negotiate in cleartext with its partner to decide whether a secure session is necessary. If both agree to a secure session, the application needs to tell AT-TLS to set up a secure session. The SIOCTTLSCTL ioctl provides the interface for the application to query or control AT-TLS.

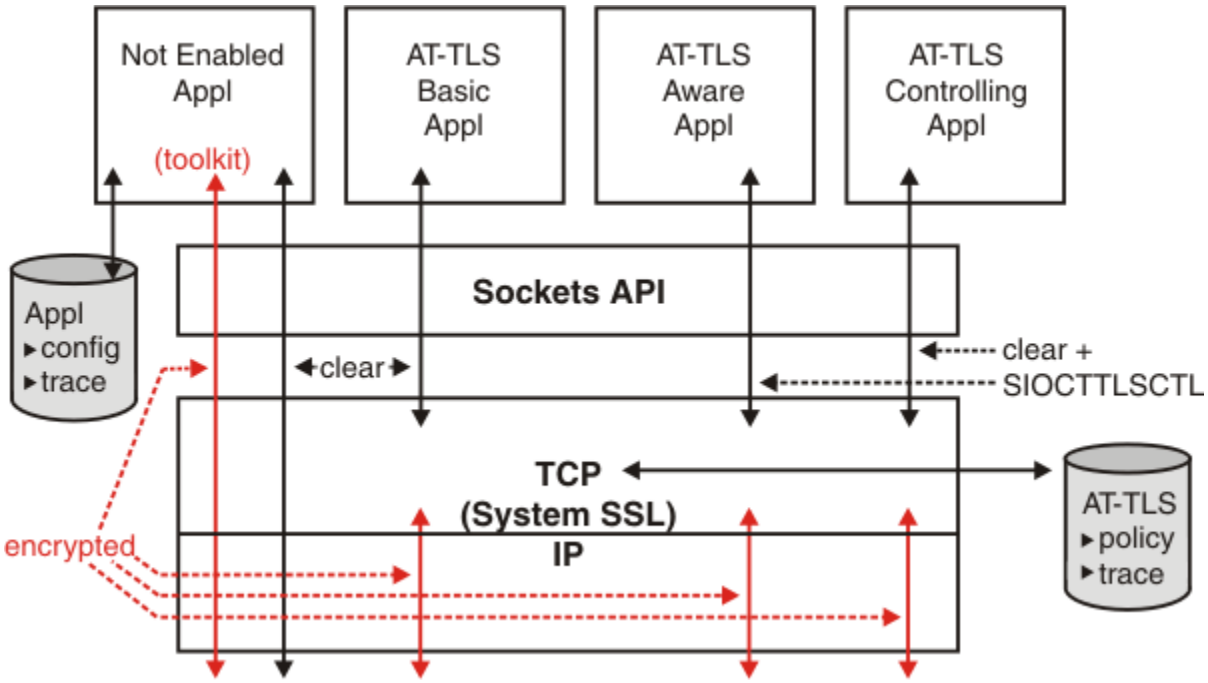


Figure 139. Application Transparent TLS

In all cases, an application using a socket enabled with AT-TLS continues to send and receive text data in the clear while encrypted data flows over the network. This allows the use of TLS with applications that cannot be modified or that cannot incorporate one of the available tool kits. The partner application must also support TLS protocols, either by using AT-TLS or an available TLS toolkit.

AT-TLS configuration in PROFILE.TCPIP

AT-TLS support is controlled by the TTLS or NOTTLS parameter on the TCPCONFIG statement in PROFILE.TCPIP. AT-TLS is enabled by specifying TTLS. The information required to negotiate secure

© Copyright IBM Corp. 2000, 2022

connections is provided to the stack by AT-TLS policies configured in Policy Agent. When AT-TLS is enabled and a newly established connection is first used, the TCP layer of the stack searches for a matching AT-TLS policy installed from the Policy Agent. If no policy is found, the connection is made without AT-TLS involvement.

TCP/IP stack initialization access control

A TCP/IP stack initializes before Policy Agent installs configured policies into the stack. This leaves a window of time where connections that should be covered by AT-TLS are clear text connections. The RACF resource EZB.INITSTACK.*sysname*.*tcpname* in the SERVAUTH class is used to block stack access, except for the user IDs permitted to the resource. While in this initialization window, any socket request from an unauthorized application receives the same errno (EAGAIN with JrTcpNotActive) received prior to the stack coming up.

Checking is done only if the TCP/IP profile activates AT-TLS. If there is no profile in the SERVAUTH class covering this resource name, all socket requests fail, including those from Policy Agent. Checking ceases the first time that the Policy Agent indicates AT-TLS policy is complete, or if a TCP/IP profile change deactivates AT-TLS.

When the limited access window begins, non-scrollable message EZZ4248E is written to the system console stating that TCP/IP is waiting for Policy Agent to install AT-TLS policies. The message is released when the restriction ends. You can delay the start of AUTOLOG procedures during this window of time by specifying the optional DELAYSTART parameter with the TTLS subparameter on the AUTOLOG entry for that procedure; when specified, the procedure will start after the EZZ4248E message is deleted and message EZZ4250I is issued indicating that AT-TLS services are available.

You must permit a limited set of administrative applications to the profile to ensure full initialization of the stack. If Policy Agent is dependent on other applications in your environment, they must also be permitted. You can permit other applications that do not require AT-TLS and that you want to start prior to general applications. At a minimum, the following applications should be permitted to the profile:

- Policy Agent
- OMPROUTE
- SNMP agent and subagents
- NAMED

For examples of the security product commands needed to create this resource profile name and grant users access to it, see member EZARACF in sample data set SEZAINST.

Options for configuring AT-TLS security

AT-TLS is configured using a set of configuration statements and parameters coded into a flat file, which is parsed by the Policy Agent to establish the AT-TLS policy for each TCP/IP stack. In a complex environment, this file can become large. For this reason, there are two alternatives for creating the Policy Agent files.

Option 1: Use the IBM Configuration Assistant for z/OS Communications Server

The IBM Configuration Assistant for z/OS Communications Server, an optional GUI-based tool, provides a guided interface for configuring TCP/IP policy-based networking functions. You can use the IBM Configuration Assistant for z/OS Communications Server to generate the Policy Agent files.

The IBM Configuration Assistant for z/OS Communications Server is a z/OS Management Facility (z/OSMF) task. z/OSMF provides a web browser interface for a variety of z/OS system management functions. When you invoke the IBM Configuration Assistant for z/OS Communications Server in z/OSMF, the IBM Configuration Assistant for z/OS Communications Server runs natively in the z/OS system and you can access it through a web browser.

Through a series of wizards and online help panels, you can use the IBM Configuration Assistant for z/OS Communications Server to create AT-TLS configuration files for any number of z/OS images with any number of TCP/IP stacks per image. Using the IBM Configuration Assistant for z/OS Communications Server, there are three types of reusable objects:

- Traffic descriptors that define the local application, by describing the TCP traffic with ports or identifying the application using its jobname.
- Security levels that define the different ways to protect data, such as the encryption level.
- Requirement maps that map traffic descriptors to security levels. A single requirement map should contain a complete set of security requirements that will govern the level of security for multiple IP traffic types.

For each TCP/IP stack, you create a set of connectivity rules that indicate the data endpoints and indicate which requirement map will govern security between the data endpoints.

The IBM Configuration Assistant for z/OS Communications Server comes with a number of IBM-supplied traffic descriptors, security levels, and requirement maps that are easily applied, or you can use the IBM-supplied definitions as the basis for your own set of reusable objects.

The IBM Configuration Assistant for z/OS Communications Server can dramatically reduce the amount of time that is required to create AT-TLS policy files, contributing to ease of configuration and maintenance. Because of the inherently complex nature of z/OS security, using the GUI can help you ensure that you have a consistent and easily manageable interface for implementing AT-TLS security.

This information primarily describes option 2, manual configuration. However, if you are using the IBM Configuration Assistant for z/OS Communications Server, reading this information will help you understand security concepts and the relationship between Policy Agent and AT-TLS function.

Option 2: Manual configuration

You can manually create the AT-TLS policy configuration files by coding all the required statements in a z/OS UNIX file or MVS data set. There are a large number of configuration options provided by AT-TLS policy statements that permit advanced users to carefully fine-tune AT-TLS policy on a per-stack basis. This information describes the procedure for creating an AT-TLS policy by manually creating and editing the configuration files. For details about the AT-TLS policy statements, see z/OS Communications Server: IP Configuration Reference.

Specifying the AT-TLS configuration file based on Policy Agent role

The Policy Agent can act as a policy server, a policy client, or neither. For more information on these different roles, see “Policy types and infrastructure overview” on page 787. Regardless of which option is used to configure AT-TLS policies, the resulting configuration files need to be specified using different statements, depending on the role of the Policy Agent.

- If you are using the Policy Agent as a policy client that retrieves AT-TLS policies from the policy server, specify the configuration files using the DynamicConfigPolicyLoad statement on the policy server.
- If you are using the Policy Agent as a policy client, but the policy client does not retrieve AT-TLS policies from the policy server, specify the configuration files using the TTLSConfig statement on the policy client.
- If you are not using a policy client/policy server environment, specify the configuration files using the TTLSConfig statement on the single Policy Agent.

When this information refers to configuration files, keep in mind where the files should exist, based on the role of the Policy Agent.

AT-TLS policy configuration

AT-TLS policy is provided to the stack by the Policy Agent. The Policy Agent main configuration file contains a TcpImage statement for each stack that is to receive policy, and can optionally contain a CommonTTLSConfig statement that identifies a local shared AT-TLS policy file.

The TcpImage statement identifies the z/OS UNIX file or MVS data set that contains policy for that stack. This policy file can contain a TTLSConfig statement to identify the z/OS UNIX file or MVS data set that contains the local AT-TLS policy. The TTLSConfig statement is required for each stack that is to receive AT-TLS policy. If both a TTLSConfig statement and a CommonTTLSConfig statement are defined, the specified CommonTTLSConfig file is processed before the TTLSConfig policy file specified for that stack.

On the policy server, use the DynamicConfigPolicyLoad statement to specify the remote AT-TLS policies. On the policy client, use the PolicyServer statement to retrieve the remote AT-TLS policies from the policy server.

Within the AT-TLS policy file, AT-TLS rules define a set of conditions that are compared to connections when policy is mapped during connect, or at the first select for readable or writable, poll for readable or writable, send, receive, or SIOCTTLSCTL ioctl. If a rule match is found, AT-TLS transparently provides TLS protocol control for the connection based on the security attributes specified in the actions associated with the rule.

AT-TLS rules

A TTLSRule statement consists of a set of conditions that are compared against the connection being checked. When a match is found, policy lookup stops and the connection is assigned the actions associated with the rule. The rule conditions are:

- LocalAddr - Local IP address or addresses
- RemoteAddr - Remote IP address or addresses
- LocalPortRange - Local port or ports
- RemotePortRange - Remote port or ports
- Jobname - Job name of the owning application or wildcard job name
- Userid - User ID of the owning process or wildcard user ID
- Direction - Inbound if applied to a passive socket (established by accept), Outbound if applied to an active socket (established by connect), or Both

Direction and at least one other condition must be specified. Other rule considerations include:

- If a condition is not specified, that condition is not considered when comparing the rule and the connection for a match.
- Multiple values can be specified for the IP address and port conditions, either directly in the condition or as a referenced group.
- IPv6 addresses are valid in all environments.

Each TTLSRule statement can also have a priority. Priority values can be integers in the range 1 - 2000000000, with 2000000000 being the highest priority. When assigning priorities, you should skip some values to allow for future rule insertion between existing rules. Policy Agent orders rules in alphabetical order within priority.

Tip: If connections can map to more than one rule, always use priority and leave priority space between rules.

AT-TLS actions

A TTLSRule statement can reference up to three actions:

- The TTLSGroupActionRef parameter includes the name of a globally defined TTLSGroupAction statement
- The TTLSEnvironmentActionRef parameter includes the name of a globally defined TTLSEnvironmentAction statement
- The TTLSConnectionActionRef parameter includes the name of a globally defined TTLSConnectionAction statement

The TTLSGroupActionRef parameter is required, and the TTLSGroupAction statement must specify TTLSEnabled ON or TTLSEnabled OFF. If TTLSEnabled OFF is specified, no additional specifications are needed. Otherwise, the AT-TLS environment action is required and the AT-TLS connection action is optional. Each action represents a scope of control.

When an AT-TLS action statement is deleted or replaced, it is considered stale. New connections will not map to a stale action. Connections that mapped to an action that later becomes stale continue to use the resources associated with the stale action until the connection closes.

AT-TLS group action

This action defines whether AT-TLS is enabled, allows trace settings, and provides the ability to set unique Language Environment variables for the Language Environment process that will be started for the action. Many TTLSRule statements can reference the same TTLSGroupAction statement. In a simple implementation of AT-TLS, you need to specify only TTLSEnabled ON on the TTLSGroupAction statement.

The AT-TLS group action represents a single Language Environment process and enclave, and initializes one instance of the System SSL DLL. Global attributes are owned by this action. Each AT-TLS group has a main task, a logging task, and a dynamic pool of pthreads that handle System SSL secure environment and secure connection management. When a stale group has no remaining connections, the pthreads, tasks, and Language Environment process are removed.

Guideline: Use as few AT-TLS group actions as necessary.

AT-TLS environment action

This action requires a key ring name (either RACF or gskkyman format), and the handshake role (client or server) this half of the connection will assume. Cipher suites and trace settings can also be set. There are several advanced parameters available if needed, but in a simple implementation of AT-TLS, you need to specify only a key ring and the handshake role.

The AT-TLS environment action is used to create System SSL environments. A key ring and SSL Session ID (SID) cache are examples of attributes owned by a System SSL environment. The AT-TLS environment action initializes a System SSL environment within the Language Environment process that was created to represent an AT-TLS group action. Several System SSL environments can exist within a single AT-TLS group. The same TTLSEnvironmentAction statement can be used to create similar System SSL environments in the same or different groups. AT-TLS dynamically creates instances of System SSL environments as needed. AT-TLS deletes System SSL environments when they have had no connections using them for a period of ten to twenty minutes. If the TTLSEnvironmentAction statement used to create a System SSL environment becomes stale, AT-TLS deletes the environment when it has no remaining connections. Connections associated with the same server application or same client user ID can share a System SSL environment. Connections that share an existing System SSL environment avoid the processing required to initialize an environment, such as opening a key ring. Connections between the same partners can also reuse recent session information in the SID cache, allowing them to use the SSL short handshake that requires less processing. System SSL connection resources are released when the connection closes.

AT-TLS connection action

The AT-TLS connection action represents attributes at the connection level. This action is optional, and is needed only when a subset of connections within an AT-TLS environment must have different parameters. Handshake role, security version, cipher suites, and tracing are examples of attributes that can be changed at the connection level. In a simple implementation of AT-TLS, you do not need to specify this action.

System SSL connections are initialized within a System SSL environment. Use the AT-TLS connection action to override attributes specified at the SSL environment layer. System SSL connection resources are released when the connection closes.

Getting started with AT-TLS

Assume you have a TCP client and server application pair running on z/OS platforms. This application handles sensitive data, and you want this application to be used only with the TLS protocol. The server application runs under the job name of XYZSRV, and creates a passive TCP socket bound to IP address INADDR_ANY and port 5000. The client application runs as a command, issued by TSO or z/OS UNIX interactive users, and connects to port 5000.

To complete AT-TLS security setup for this sample environment, you need to create both server and client key rings. The server key ring needs to contain a server certificate, and any certificates used to sign it. The server needs access to the private keys of the server certificate. The client key ring needs the root certificate used to sign the server certificates. For a TLS/SSL primer and some step-by-step examples, see Appendix B, “TLS/SSL security,” on page 1329. For more information on managing key rings and certificates with RACF and the RACDCERT command, see z/OS Security Server RACF Security Administrator's Guide. For detailed information on managing key rings and certificates with gskkyman, see z/OS Cryptographic Services System SSL Programming.

Configuring the server system

On each z/OS system where you run the server application, see Table 58 on page 1126 for the tasks needed to configure the server.

Table 58. AT-TLS configuration for the server system

Task	Specification
Create key ring	Create server key ring with server certificate and necessary certificate authority certificates.
Create Policy Agent files	1. Create a Policy Agent main configuration file containing a TcpImage statement for the server stack. 2. Create a Policy Agent image configuration file for the server stack. 3. If AT-TLS policies are to be retrieved from the policy server, create image-specific AT-TLS configuration files, and optionally, common AT-TLS configuration files, on the policy server.

Table 58. AT-TLS configuration for the server system (continued)

Task	Specification

Add AT-TLS configuration	1. For local AT-TLS policies, add a TTLSConfig statement to the Policy Agent image configuration file, identifying the TTLSConfig policy file location:
	`TTLSConfig  serverpath`
	2. For remote AT-TLS policies, add a PolicyServer statement to the policy client image configuration file:
	`PolicyServer`
	`{`
	`  ClientName  name`
	`  PolicyType  TTLS`
	`  {`
	`  …`
	`  }`
	`  …`
	`}`
	Add a DynamicConfigPolicyLoad statement to the policy server main configuration file:
	`DynamicConfigPolicyLoad  clientname`
	`{`
	`  PolicyType TTLS`
	`  {`
	`  PolicyLoad  serverpath`
	`  }`
	`  …`
	`}`

Add statements to the AT-TLS policy file	Add the AT-TLS policy statements to the *serverpath* file:
	`TTLSRule  XYZServerRule`
	`{`
	`  LocalPortRange  5000`
	`  JobName  XYZSRV`
	`  Direction  Inbound`
	`  TTLSGroupActionRef  XYZGroup`
	`  TTLSEnvironmentActionRef  XYZServerEnvironment`
	`}`
	`TTLSGroupAction  XYZGroup`
	`{`
	`  TTLSEnabled  On`
	`}`
	`TTLSEnvironmentAction  XYZServerEnvironment`
	`{`
	`  TTLSKeyRingParms`
	`  {`
	`# Assuming SERVER is the owner of the keyring server_key_ring`
	`  Keyring  SERVER/server_key_ring`
	`  }`
	`  HandshakeRole  SERVER`
	`  Trace  7`
	`}`

Set up InitStack access control	1. Define the EZB.INITSTACK.*sysname*.*tcpname* profile for each AT-TLS stack. 2. Permit administrative applications to use the stack before AT-TLS is initialized. For examples of the security product commands needed to create this resource profile name and grant users access to it, see member EZARACF in sample data set SEZAINST.
Enable AT-TLS	Set TCPCONFIG TTLS in PROFILE.TCPIP.

Configuring the client systems

On each z/OS system where you run the client application, see Table 59 on page 1128 for the tasks needed to configure the client.

Table 59. AT-TLS configuration for the client system

Task	Specification
Create key ring	Create/specify a key ring for each client: • **When using TLS server authentication without client authentication**, there are two approaches for defining your client-side key rings. As you decide between these approaches it is important to know that a complete copy of a key ring's contents will be loaded into TCP/IP above-the-bar private memory for each user that opens a connection under the protection of an AT-TLS client rule that uses that key ring. If the key ring in question contains a lot of certificates and a non-trivial number of users will use the key ring, it could consume a large amount of above-the-bar TCP/IP private memory. 1. Specify the virtual CERTAUTH key ring (`*AUTH*/*`) for the client AT-TLS rule. This virtual key ring contains every trusted certificate authority (CA) certificate in the RACF database, thus eliminating the need to create client-specific key rings with the required CA certificates. 2. Create a client key ring for each client user ID using the specified key ring name that only contains the necessary CA certificates. Due to the way the key ring name is specified in the client-side AT-TLS policy (defined in a later step), when it is time to open the key ring,, System SSL looks for a key ring of the specified name that is owned by the user ID under which the client is executing. While requiring more administration than the virtual key ring, this approach could be a good alternative if there are a large number of trusted CA certificates in the RACF database. • **When using TLS server authentication with client authentication**, create a client key ring for each client user ID using the specified key ring name. This key ring must contain the client certificate and private key, the CA certificates in the client's trust chain, as well as any CA certificates required to validate the server's certificate. Due to the way the key ring name is specified in the client-side AT-TLS policy (defined in a later step), when it is time to open the key ring, System SSL looks for a key ring of the specified name that is owned by the user ID under which the client is executing.
Create Policy Agent files	1. Create a Policy Agent main configuration file containing a TcpImage statement for the client stack. 2. Create a Policy Agent image configuration file for the client stack. 3. If AT-TLS policies are to be retrieved from the policy server, create image-specific AT-TLS configuration files, and optionally, common AT-TLS configuration files, on the policy server.

Table 59. AT-TLS configuration for the client system (continued)

Task	Specification

Add AT-TLS configuration	1. For local AT-TLS policies, add a TTLSConfig statement to the Policy Agent image configuration file, identifying the TTLSConfig policy file location:
	`TTLSConfig  clientpath`
	2. For remote AT-TLS policies, add a PolicyServer statement to the policy client image configuration file:
	`PolicyServer`
	`{`
	`  ClientName  name`
	`  PolicyType  TTLS`
	`  {`
	`  …`
	`  }`
	`  …`
	`}`
	Add a DynamicConfigPolicyLoad statement to the policy server main configuration file:
	`DynamicConfigPolicyLoad  clientname`
	`{`
	`  PolicyType TTLS`
	`  {`
	`  PolicyLoad  clientpath`
	`  }`
	`  …`
	`}`

Add statements to the AT-TLS policy file	Add the AT-TLS policy statements to the *clientpath* file:
	`TTLSRule  XYZClientRule`
	`{`
	`  RemotePortRange  5000`
	`  Direction  Outbound`
	`  TTLSGroupActionRef  XYZGroup`
	`  TTLSEnvironmentActionRef  XYZClientEnvironment`
	`}`
	`TTLSGroupAction  XYZGroup`
	`{`
	`  TTLSEnabled  On`
	`}`
	`TTLSEnvironmentAction  XYZClientEnvironment`
	`{`
	`  TTLSKeyRingParms`
	`  {`
	`# The Keyring value is not qualified with the key ring owner.`
	`# This allows the rule to be used by multiple clients with`
	`# different user IDs.`
	`# For rules that are only used under one specific client user ID, the user ID`
	`# of the key ring owner can be specified in the Keyring parameter`
	`# as USERID/client_key_ring.`
	`# Specifying the owning user ID can simplify diagnostics.`
	`  Keyring  client_key_ring`
	`  }`
	`  HandshakeRole  CLIENT`
	`  Trace  7`
	`}`

Table 59. AT-TLS configuration for the client system (continued)	
Task	**Specification**
Set up InitStack access control	1. Define the EZB.INITSTACK.*sysname*.*tcpname* profile for each AT-TLS stack. 2. Permit administrative applications to use the stack before AT-TLS is initialized. For examples of the security product commands needed to create this resource profile name and grant users access to it, see member EZARACF in sample data set SEZAINST.
Enable AT-TLS	Set TCPCONFIG TTLS in PROFILE.TCPIP.

Steps for starting AT-TLS and verifying its operation

After configuring the server and client systems, use these steps to start and verify AT-TLS.

Before you begin

- Perform the tasks in AT-TLS configuration for the server system and AT-TLS configuration for the client system.
- Review your syslogd configuration to verify that messages written by Policy Agent and TCP/IP stacks are saved in the wanted files. AT-TLS syslogd messages are written to the daemon facility by default.
- Start syslogd.

You are now ready to start the sample AT-TLS environment and verify its operation.

Procedure

Perform the following steps to start AT-TLS and verify its operation:

1. Start the TCP/IP stacks.
2. Start the administrative applications required to successfully run Policy Agent, such as OMPROUTE and LDAP.
3. If System SSL needs to access Integrated Cryptographic Services Facility (ICSF), start ICSF. For information about using cryptographic features with System SSL, see z/OS Cryptographic Services System SSL Programming.
4. Start Policy Agent on all participating systems and verify that there were no policy errors in processing the policy files.
5. Verify that the participating TCP/IP stacks have received AT-TLS policy and released console message EZZ4248E.
6. Start server application and verify that it starts without errors.
7. Start client applications. Review the AT-TLS trace messages in the syslogd output on both the client and server systems. Verify that connections are mapping to the intended policy and no handshake errors occur. The info messages `EZD1281I TTLS Map` and `EZD1283I TTLS Initial Handshake` show the policy used and result of TLS handshake negotiation. The error message `EZD1286I TTLS Error` shows any failures.

Results

For information on common AT-TLS startup errors, see z/OS Communications Server: IP Diagnosis Guide.

Application compatibility with AT-TLS

Most applications can use AT-TLS. However, some applications should not be configured to use AT-TLS. Any application that is already configured to use SSL or TLS protocols should not use AT-TLS. Use of AT-TLS would result in encrypting data that is already encrypted. The receiving partner would not be able to decipher the data that had been encrypted twice. If the application can be configured to use clear text or the application uses the SIOCTTLSCTL ioctl, AT-TLS can provide support.

The only TCP/IP application supplied with z/OS Communications Server that already supports the TLS protocol is the FTP client, which provides two modes of TLS support. One uses native calls to System SSL and the other uses AT-TLS. To configure the FTP client to use AT-TLS, code the TLSMECHANISM ATTLS statement in FTP.DATA. For more information, see "Steps for migrating the FTP client to use AT-TLS" on page 703.

AT-TLS does not support web servers using the Fast Response Cache Accelerator (FRCA) support in TCP/IP. AT-TLS ignores policy for connections using FRCA. The sockets are treated as if they did not match any AT-TLS rules.

AT-TLS does not support applications that use the Pascal sockets API. AT-TLS ignores all Pascal sockets. The sockets are treated as if they did not match any AT-TLS rules. TCP/IP applications that use the Pascal API include:

- TSO TN3270E Telnet client
- LPD server

Policy considerations

Policy is a powerful tool for configuring and managing your applications that use AT-TLS. Before configuring the AT-TLS policy, review the general syntax rules in the Policy Agent and policy applications topic in z/OS Communications Server: IP Configuration Reference.

Reusable objects

With several of the AT-TLS rule conditions and action parameters, you can reference named objects. If you are going to use the same definition in several rules or actions, it is easier to create a single named object and refer to it, rather than repeating the definition. This also makes changing these definitions easier and more accurate. For example, the IpAddrGroup statement can be used to identify groups of IP addresses that are used in several rules. You might find it useful to define TTLSRule statements that reference an IP address group with a name, such as LocalHost. In each stack's AT-TLS policy file, you would define that IP address group with all of the addresses local to that stack. That single reference can then be used throughout the policy to easily represent all local IP addresses, without re-coding the local addresses in each rule condition.

Common AT-TLS configuration file

The common AT-TLS configuration file should contain all of the policy that is common to multiple stacks. When the policy agent reads a policy file for a given stack, the contents of the common AT-TLS configuration file are logically added before the contents of the stack-specific file. Rules and actions in the common AT-TLS configuration file can reference objects, such as a local IpAddrGroup statement, that are defined in the stack-specific AT-TLS configuration file. Rules and actions in the stack-specific file can also reference objects that are defined in the common file.

If Policy Agent encounters multiple objects of the same type and name, the last occurrence is the one that is used. You can take advantage of this by defining an object that is used on many stacks in the common file, and then overriding it in the stack-specific file of a specific stack.

Exempting specific connections from AT-TLS

In some cases, you might want to have the majority of users of an application using AT-TLS, but then exempt a small group of users. You might find that it is simpler to define a broad AT-TLS policy for the

application, and then define a higher priority rule for the exempt users. The simplest way to do this is to define a TTLSGroupAction statement with the TTLSEnabled parameter set to OFF. The exempt user rule would then reference this action. When the TTLSGroupAction statement specifies the TTLSEnabled parameter as OFF, the TTLSEnvironmentActionRef parameter on the corresponding rule is not required.

Action refresh

When Policy Agent is stopped and restarted, or when policy files are changed, policy objects that are currently in use might be deleted or replaced. When an AT-TLS action is deleted or replaced, connections using the old object continue processing without change. Connections that search AT-TLS policy after the change use the new action objects. A change to an AT-TLS group action causes a new Language Environment process to be created, along with new SSL environments for each user or application environment associated with that group. A change to an AT-TLS environment action causes new SSL environments to be initialized within each group with which that environment is associated. System SSL reopens key rings and certificates, and creates an empty session ID cache when it initializes an SSL environment.

There are cases when a change is made that is not reflected by a change in the action. For example, the default certificate in a key ring might change. The key ring name has not changed, but there is a need to open a new environment. Simply refreshing policy will not refresh the AT-TLS environment action in AT-TLS, because no values within the action have changed. To force a refresh in AT-TLS, some parameter must be changed. The EnvironmentUserInstance parameter can be used for this purpose. Incrementing the instance number forces a refresh of AT-TLS without changing any of the security parameters. Similarly, changes to the contents of the environment file named in a group action will not be applied until the group action is changed. The GroupUserInstance parameter can be used to force an AT-TLS refresh of the group, creating a new Language Environment process using the new environment file contents.

Sometimes after you have made a change to an AT-TLS policy, the changed policies are not automatically reinstalled by the Policy Agent; new connections might fail until the policies are reinstalled. If you see AT-TLS connection setup errors with message EZD1286I or EZD1287I after you made an AT-TLS configuration change, you can force all AT-TLS policies to be reinstalled by refreshing the Policy Agent. From the MVS console, issue the MODIFY *procname*,REFRESH command. For more information about controlling the refresh of polices using the TcpImage and PEPInstance statements, see z/OS Communications Server: IP Configuration Reference.

Achieving the basic level of security

To achieve the basic level of security, take the following actions:

1. Pick the handshake roles.
2. Specify the key ring.

Picking the handshake roles

Every secure connection must have one end using the HandshakeRole Client, and the other end using the HandShakeRole Server or HandShakeRole ServerWithClientAuth.

In the SSL and TLS protocols, the client side sends a ClientHello handshake record and the server side responds with a ServerHello handshake record. If both sides send ClientHello records, the handshake fails. If both sides wait for a ClientHello, the handshake times out.

The role played in the handshake is independent of whether the application is used as the client, server, or peer. It is also independent of which application does the listen() and accept() and which does the connect(). The primary consideration for deciding which role an application should play is certificate management.

- An application designated as HandshakeRole Server must have a user or site certificate on its key ring. It must have access to the private keys of its user or site certificate. It might also need other certificates on its key ring, required for authentication.

The partner application designated as HandshakeRole Client must have a key ring that contains the root certificates required to authenticate the server's certificate. The client does not need access to the private keys of any certificate. Clients can share this key ring.

- An application designated as HandshakeRole ServerWithClientAuth might need additional root certificates on its key ring to authenticate client certificates. The server decides whether a client must present a user certificate and what constitutes an acceptable certificate.

  The partner application designated as HandshakeRole Client decides whether to present a user certificate when challenged. To present a user certificate, it must have a private key ring with that user certificate. It must have access to the private keys of the user certificate. The client key ring must also contain the root certificates required to authenticate the server's certificate.

When the HandshakeRole parameter is set to ServerWithClientAuth, a certificate request is sent to the client during the handshake. The client can send its certificate to the server, which can then validate the certificate. The level of validation done by the server is controlled by the setting of the ClientAuthType parameter. The following chart summarizes the differences between the parameter settings:

Table 60. ClientAuthType parameter settings

ClientAuthType	**Client certificate required or optional**	**Certificate validation**
PassThru	Optional	None
Full	Optional	Certificate is validated against keyring, if provided
Required	Required	Certificate is validated against keyring
SAFCheck	Required	Certificate is validated against keyring and must be associated with a user ID in the security product

If the certificate fails validation, the secure connection does not initialize successfully. The default setting for ClientAuthType is Required. For applications that do not issue AT-TLS ioctl calls to obtain the certificate or user ID of the client, the Required setting ensures that any client that connects provides a valid client certificate. The ClientAuthType setting of PassThru should be used only for applications that will get the certificate from AT-TLS using the SIOCTTLSCTL ioctl call and implement their own security methods to validate the certificate. The ClientAuthType setting of Full can be used for applications that want to validate the client certificate if provided, but do not require a client certificate to establish a secure connection. The ClientAuthType setting of SAFCheck provides an additional level of security when using client authentication by requiring AT-TLS to derive a user ID associated with the certificate. An application can then get the user ID from AT-TLS using the SIOCTTLSCTL ioctl after the secure connection has been established. The user ID can be used to either verify the user ID presented by the client during the application's protocol flows or eliminate the need for a user ID to be sent by the client.

Specifying the key ring

A fully-qualified SAF key ring name is specified as *userid/keyring*. A value that begins with a forward slash is interpreted as the path name of a gskkyman key ring. A PKCS #11 token name is specified as *TOKEN*/*token-name*. *TOKEN* indicates that the specified key ring is actually a token name.

When creating AT-TLS rules for cases where the protected connections will always use the same SAF keyring, owned by the same user ID, you should specify the fully qualified key ring name in the form *userid/keyring* on the Keyring parameter. Coding the user ID on the Keyring parameter can be helpful when diagnosing TLS handshake issues. However, for cases where multiple TLS client user IDs each have their own key ring using a common keyring name, you can omit the userid and forward slash so that only the keyring name itself is specified. When the user ID is omitted, AT-TLS and System SSL assume the currently executing user ID owns the keyring when the TLS handshake takes place. This allows multiple

clients that use a common key ring name to share the same AT-TLS rule. Note that doing this, however, adds diagnostic steps to determine which key ring was being used in case an error occurs.

Configuring more sophisticated security

This topic describes decisions you can make to achieve more elaborate levels of security.

Cipher suite specification

You can specify the exact list of TLS cipher suites to be allowed for TLS sessions. You should not include any cipher suites that you do not want to allow. The order of the cipher suite list is important, as it determines the priority order for selecting a cipher suite during a TLS handshake. In general, System SSL selects cipher suites according to the server's order of usage preference. The first cipher suite in the server's list that is present in the client's list and is also supported for the selected TLS protocol version is selected. Other implementations might work differently.

AT-TLS does not pass any cipher suites to System SSL by default. For the list of cipher suites supported and the default order used if none is specified, see z/OS Cryptographic Services System SSL Programming.

Guideline: Code TTLSCipherParms statements to support newer cipher algorithms, such as elliptical curve cipher suites, AES Galois Counter Mode (GCM) cipher suites, or cipher suites that use SHA2-based digests.

Tip: If you configure your AT-TLS policies using the Network Configuration Assistant (NCA), you can select a recommended cipher suite list.

Requirement: Integrated Cryptographic Service Facility (ICSF) must be active for elliptic curve ciphers and for ciphers using AES-GCM or Cha-Cha Poly algorithms.

If the CSFSERV class is defined, the user ID that is associated with the AT-TLS application must have READ access to the following resources in that class:

- CSF1DVK
- CSF1GAV
- CSF1GKP
- CSF1PKS
- CSF1PKV
- CSF1SKD
- CSF1SKE
- CSF1TRC
- CSF1TRD
- CSFOWH

For more information about elliptic curve cryptography support, see z/OS Cryptographic Services System SSL Programming.

Protocol versions

TLSv1.2 and TLSv1.3 provide stronger security than SSLv2, SSLv3, TLSv1.0 and TLSv1.1. By default, SSLv2 and SSLv3 are not supported in AT-TLS policy. To use older applications that do not support the newer protocols, set SSLv2 or SSLv3 on.

If you plan to use the TLSv1.3 protocol, you must ensure that you specify at least one TLSv1.3 cipher suite.

Using TLSv1.3 protocol support

The TLS Version 1.3 protocol is a major revision to the TLS protocol that is intended to provide better security and improve handshake performance. To use TLSv1.3, there are certain configuration requirements:

- Cipher specifications: The cipher specifications that are valid for TLSv1.2 and earlier protocols are not supported for TLSv1.3. AT-TLS supports three TLSv1.3 cipher suites: TLS_AES_128_GCM_SHA256, TLS_AES_256_GCM_SHA384, and TLS_CHACHA_POLY1305_SHA256. For an AT-TLS rule that enables TLSv1.3, you must specify one or more of these cipher suites.

  Guideline: When multiple TLS protocol versions are supported, it is recommended that you include the TLSv1.3 cipher specifications at the beginning of the cipher suite list.
- Supported groups or curves for certificate selection (TLSv1.0 and later): A supported groups or curves list is configured for the client with the ClientECurves parameter. It is sent to the server in a supported groups extension. The client's list is used by the server to select a group or curve for use during the key exchange process and to guide selection of the server's certificate when elliptic curve-based certificates are used.
- Key share groups: TLSv1.3 uses client and server key shares to facilitate the encryption of TLSv1.3 handshake messages and to determine the key exchange algorithms. When using TLSv1.3, each TLS partner must provide its key share.

  Key share group settings are configured in AT-TLS policy. For a client, the key share groups or curves for which a key share should be generated are specified with the ClientKeyShareGroups parameter. The client generates a key share for each group in the ClientKeyShareGroups list that is also in the ClientECurves list. The order of the key shares in the key_share extension sent to the server is based on the ClientECurves order.

  For a server, the supported key share groups or curves are specified with the ServerKeyShareGroups parameter. When the server receives the client's key_share extension, the server will use the client's preference order in selecting a group that is also supported for the server. The selected group is used to encrypt the remaining portions of the TLSv1. 3 handshake.
- Signature and hash algorithms: The list of signature and hash algorithms supported by the TLS client and server are specified in order of preference using the SignaturePairs parameter. These algorithms are used in the TLSv1.3 handshake for digital signatures of X.509 certificates and TLS handshake messages. The client sends its supported list to the server in the supported groups extension to guide the server's certificate selection and the selection of an algorithm for use in protecting TLSv1.3 handshake messages. If the server is enabled for client authentication, the server sends its supported list to the client in the supported groups extension for the same type of guidance.

  Tip: When different algorithm pairs need to be specified for the key exchange and certificate selection, SignaturePairsCert can be configured to guide the certificate selection.

  Rule: When using RSA certificates for TLSv1.3 handshakes, SignaturePairs should include the RSASSA-PSS signature algorithms for generating the signatures for TLSv1.3 handshake messages.
- Middlebox compatibility mode: When TLSv1.3 was under development by the IETF, there were issues with some network-based devices that act as TLS proxies. These devices, called "middleboxes" were not able to properly parse pure TLSv1.3 handshake messages. To accommodate these middleboxes, RFC 8446 (TLSv1.3 specification) defined a compatibility mode. The AT-TLS parameter MiddleBoxCompatMode can be set on to enable compatibility mode. It is disabled by default.

For more details on TLSv1.3 selection of signature algorithm pairs, groups and curves, and X.509 certificates and other considerations, see z/OS Cryptographic Services System SSL Programming.

Certificate validation

You can specify that certificates should be validated by using the following methods:

- Only the method described in RFC 2459
- Only the method described in RFC 3280

- Only the method described in RFC 5280
- Any of the methods that are described in RFC2459, RFC3280 and RFC5280

If 128-bit minimum or 192-bit minimum Suite B profiles are configured, the certificates will be validated as described in RFC 5280 and RFC 5759.

FIPS 140-2 support

You can configure AT-TLS to support FIPS 140-2. Specify either 0n, Level1, Level2, or Level3 for the FIPS140 statement of the TTLSGroupAction statement.

For information about configuring System SSL to run in FIPS 140-2 mode, see the System SSL and FIPS 140-2 topic in z/OS Cryptographic Services System SSL Programming.

Restriction: The FIPS 140-2 standard does not define support for TLSv1.3 or the new cipher suites defined for it. Enabling both the TLSv1.3 protocol and FIPS support will result in an error.

Requirement: ICSF must be active before starting AT-TLS groups configured to support FIPS140. For information about configuring ICSF to support FIPS 140-2, see the topic about operating in compliance with FIPS 140-2 in z/OS Cryptographic Services ICSF Writing PKCS #11 Applications.

If the RACF CSFSERV class is defined, the user ID associated with the TCP/IP stack and any application user ID that is using the TTLSGroup must be given READ access to the CSFRNG resource within the CSFSERV class. If the CSFSERV class is defined and Diffie-Hellman is being used, the application user ID must be given READ access to the CSF1TRC, CSF1DVK, CSF1GKP, CSF1GSK, CSF1GAV, and CSF1TRD resources within the CSFSERV class.

Certificate revocation support

Applications requiring validation of the partner's certificate can optionally check to see if the certificate has been revoked. Certificate revocation checking can be done using a certificate revocation list (CRL) obtained from a LDAP or HTTP server or by using the certificate revocation status obtained from an OCSP responder. You can configure any combination of certificate revocation by using the following AT-TLS policy statements:

- TTLSGskHttpCdpParms
- TTLSGskLdapParms
- TTLSGskOcspParms

Guideline: Connections that are used by System SSL to contact the CRL service should not fall under an enabled AT-TLS policy because these connections can be made before AT-TLS policy is installed.

Encryption key refresh

All of the TLS protocol versions except SSLv2 support some form of dynamic key refresh (also call "cipher reset"). The SSLv3, TLSv1, TLSv1.1, and TLSv1.2 protocols allow the encryption key to be renegotiated during a secure connection. TLSv1.3 supports dynamic key updates without renegotiation. Refreshing the encryption key provides a higher level of security for long-running connections. The AT-TLS default is to not reset the cipher.

To enable cipher reset through AT-TLS, you can specify a time interval in the range 1–1440 minutes using the ResetCipherTimer parameter. The cipher reset is requested when the timer expires and the next application read or write is completed. The time interval is restarted when the cipher has been changed. The specific behavior for a cipher reset depends on the TLS protocol version of the connection:

- For TLSv1.2 and prior versions, both ends of the secure connection must agree to perform another handshake to renegotiate the cipher. By default, both the client and server must support RFC 5746 renegotiation. The `HandshakeRole Client` end must initiate this handshake. The `HandshakeRole Server` end can send an alert to the client requesting another handshake. The client is free to ignore or postpone the request. The server is free to refuse a handshake request sent by the client for another handshake.

- For TLSv1.3, either TLS endpoint can initiate a key update for the traffic it encrypts and sends to the partner, which typically results in the partner sending a key update for the traffic in the opposite direction in response.

Additional security customization considerations

This topic provides additional information that you might want to use to customize security for your environment.

Handshake timer

Certain configuration or application protocol mismatches can lead to stalled connections. The TLS handshake protocol always expects the end of the connection configured as HandshakeRole Client to send the first message. One connection stall scenario results when both ends of the connection are configured as HandshakeRole Server and wait for a ClientHello record from the other end. Another connection stall scenario results when the end of the connection that normally sends the first application data is configured as HandshakeRole Server, but the partner application is not configured to use a secure connection. The HandshakeRole Server end is waiting for a ClientHello, and the nonsecure end is waiting for application data.

You can use the HandshakeTimeout action parameter to control the time that AT-TLS should wait during TLS handshake negotiation before resetting the connection. AT-TLS times two different handshake intervals, handshake start and handshake completion. The handshake start interval is intended to detect configuration problems that result in neither partner sending data. The handshake completion interval is intended to detect problems that might stall a handshake in one of the TLS protocol implementations.

The handshake start interval begins when AT-TLS is ready to begin a TLS handshake, and ends when the hello handshake record is received from the partner. On the initiating or active side of the connection, the handshake start interval used is five times the specified HandshakeTimeout value, because it includes:

- The network time for the ClientHello record to reach the partner if HandshakeRole is Client.
- The time a connection spends on the partner's listen backlog.
- The time before the partner causes the connection to be mapped.
- The time spent on the partner AT-TLS work queue.
- The time spent by the partner initializing a new System SSL environment, if necessary.
- The network time for the partner's ServerHello or ClientHello record to be returned.

On the listening or passive side of the connection, the handshake start interval used is the specified HandshakeTimeout value, because it includes only the network time for one or both hello records. Handshake start interval timeouts result in AT-TLS return code 5004 and a connection reset.

The handshake completion interval begins when the hello handshake record is received from the partner, and ends when the System SSL gsk_secure_connection_init() service returns to AT-TLS. The handshake completion interval used is the specified HandshakeTimeout value on either active or passive connections. Handshake completion interval timeouts result in AT-TLS return code 5005 and a connection reset.

The HandshakeTimeout action parameter has a default value of 10 seconds. If you determine that you are getting handshake time-outs that are caused by network delays or application workload rather than configuration or application errors, you should increase the value. On the other hand, if you determine that handshakes normally complete much faster in your environment and you would like to detect the occasional incorrectly configured partner more quickly, you can decrease the value.

Diagnostic traces

In addition to the steps for diagnosing AT-TLS problems described in z/OS Communications Server: IP Diagnosis Guide, you might need to collect a System SSL trace when you are diagnosing an AT-TLS problem. The only method for collecting this trace is by using GSKSRVR CTRACE, as described in z/OS Cryptographic Services System SSL Programming. You cannot use the GSK_TRACE environment variable

because it causes an abend if it is used with AT-TLS. When you use GSKSRVR CTRACE to diagnose AT-TLS problems, the job name that you specify on the JOBNAME parameter of the CTRACE command should be the TCP/IP job name rather than the application job name.

If you are not using any other features provided by the GSKSRVR started task, then you can use the sample procedure provided in the SGSKSAMP library without any changes.

Diagnosis considerations

Applications that implement SSL or TLS can control whether non-encrypted application data is included in diagnostic traces. Lower layers have access to only encrypted data. When using AT-TLS, the TCP, PFS, and SOCKAPI layers have access to non-encrypted data. The AT-TLS default is to suppress this data in CTRACE records generated by these layers to protect the application's users. If you need to see this data in these records to diagnose a problem, you can set CtraceClearText ON.

AT-TLS writes messages to syslogd using the jobname of the TCP/IP started task. The AT-TLS default behavior is to write syslogd messages to the daemon facility. Other TCP/IP functions, such as the SNMP TCP/IP subagent, also use the job name of the TCP/IP started task and specify the daemon facility name when writing records to syslogd. Because the job name and syslog facility name of the AT-TLS records and the TCP/IP function records are the same, filters cannot be used to direct the AT-TLS records to a different output file. If you want AT-TLS records to go to a different output file, configure SyslogFacility Auth on the TTLSGroupAdvancedParms statement to direct the messages from that group to the Auth facility. The job name will remain the job name of the TCP/IP started task. You can then set up filtering based on the job name of the TCP/IP started task and the auth facility in the syslogd configuration file to direct AT-TLS records to a different output file.

The Trace value is interpreted by AT-TLS as a bit map. Each of the options is assigned a value that is a power of 2. You should add together the values of each option that you want to activate.

The default Trace value is 2, which provides error messages to syslogd. While you are deploying a new policy, you might find it beneficial to specify a Trace value of 6 or 7. This provides connection info messages, in addition to error messages in syslogd. The info messages provide positive feedback that connections are mapping to the intended policy.

Trace options event (8), flow (16), and data (32) are intended primarily for diagnosing problems. Trace values larger than 7 can cause a large number of trace records to be dropped instead of being sent to syslogd.

Tip: Use a TTLSConnectionAction with a higher Trace value to diagnose problems in a production environment. You can temporarily define a high priority TTLSRule with conditions that cover only a small number of problem connections. This temporary rule can reference the same TTLSGroupAction and TTLSEnvironmentAction that your production rule references, and a TTLSConnectionAction with the Trace level you want for diagnosis.

TLS function negotiation

TLS protocols enable the TLS client and TLS server to negotiate additional functionality for a connection. If either the TLS client or TLS server does not understand a function, the function is not used on the connection. However, the TLS client or TLS server might require that the function be supported by the remote partner. If the remote partner does not support the function, the connection can be closed. Each function can be configured as Required, Optional, or Off.

- Required

  The connection ends if the remote endpoint does not accept the TLS function.

- Optional

  The function is negotiated on the connection, but the connection does not end if the remote partner does not support the function.

- Off

The function is not supported on the connection. If the remote partner requires this function, the remote partner closes this connection.

Guideline: For TLS servers, configure the functions as Optional to prevent remote partners that require this extension from being unable to connect.

Wireless performance

AT-TLS supports the following negotiated functions:

- Maximum SSL fragment length

  This TLS function negotiates the maximum size of unencrypted data that can be sent in a single SSL fragment. Without this function, 16 K is the maximum fragment length. A TLS client can negotiate a size of 512, 1 K, 2 K or 4 K. Some clients need to use the smaller size because of memory or bandwidth limitations.

- Truncated HMAC

  TruncatedHMAC is supported for TLSv1, TLSv1.1 and TLSv1.2. TLS cipher suites use the MAC construction HMAC with either MD5 or SHA-1 (RFC 2104) to authenticate record layer communications. The truncated HMAC function saves bandwidth by truncating the HMAC to 80 bits.

Certificate selection

When AT-TLS supports a server, the certificate designated as the default for the key ring is used. Use the CertificateLabel parameter to explicitly identify a different certificate that you want to use to represent the server during TLS handshakes.

If the TLS server needs to support multiple host names and multiple certificates, you can use the Server Name Indication function. The Server Name Indication function enables you to define pairs of certificate labels and host names. Use the HandshakeServerCertLabel parameter to specify these pairs for a server. To use these pairs, the TLS client must support the Server Name Indication function as well. The TLS client includes a host name during the TLS handshake, which allows the matching certificate to be used.

When AT-TLS supports a client, you can use the HandshakeServerName parameter to specify the host name to be included in the TLS handshake.

Note: For TLSv1.2 and TLSv1.3 handshakes, additional factors such as the configured signature algorithm pairs or client ecurves can also influence the selection of the server certificate. For more about these factors see “Using TLSv1.3 protocol support” on page 1135.

Session caching

The SSLv2, SSLv3, TLSv1, TLSv1.1, and TLSv1.2 protocols can cache session information based on the TLS Session ID (SID). TLS connections can request that a previous session be resumed. When session information is found in the cache, connections can use the TLS abbreviated handshake, which requires less processing.

In a similar manner, the TLSv1.3 protocol can cache session information based on a session ticket. TLS connections can request that a previous session be resumed. When session information is found in the cache, connections can use an abbreviated handshake which avoids public key encryption.

The number of SIDs or session tickets cached, the length of time that a SID or session ticket is held in the cache, and whether SIDs in the cache are available across the sysplex can be configured using the TTLSGskAdvancedParms statement. TLS Version 1.3 session tickets are not shared across the sysplex.

Session ticket generation and caching for TLS Version 1.3 can be enabled for the client and server using the TTLSGskAdvancedParms statement. The maximum size of the session ticket accepted by the client, the algorithm used by the server to encrypt the session ticket, the number of session tickets that the server sends after an initial handshake completes, and the maximum time for which a session ticket is valid for session resumption can also be configured using the TTLSGskAdvancedParms statement.

AT-TLS access control considerations

Access to key rings and certificates is verified by System SSL when SSL environments are initialized. Access to certificate private keys is verified by ICSF when asymmetric encryption services are requested that require the private keys. AT-TLS invokes System SSL services that cause these access control checks to occur on tasks created in the TCP/IP address space. TCP/IP replicates the security environment of the user running the application that owns the socket at the time AT-TLS policy is mapped, before invoking these System SSL services. As such, the access control checks use the socket application's security credentials and not those of the TCP/IP stack.

Several common application models were considered to determine the most appropriate time for replicating the security environment. Replication occurs when AT-TLS policy is mapped. Policy mapping occurs during processing of the first occurrence of connect, a SIOCTTLSCTL IOCTL, select for socket readable or writable, poll for socket readable or writable, or call that sends or receives data over the socket. This defers security environment replication for applications such as INETD until after the accept(), fork(), setuid(), and exec() sequence of services has established the server application process.

In the CICS socket environment, transaction security environments are not visible to AT-TLS support. The CICS job and all of its transactions appear to the stack as a single server application with a single z/OS UNIX process ID running in the security environment of the CICS job. All AT-TLS policy lookups, System SSL key ring authorization checks, and ICSF private key authorization checks are processed using the identity of the CICS job. Connections established, whether active or passive, can perform TLS handshake processing as either the client or server. All of the connections established by a single CICS job are able to share the session ID or session ticket cache in the SSL environment. The CICS job should use a private key ring with a server certificate. The key ring used must contain the chain of root certificates needed to validate the server certificate it presents to the client. If the server requires client authentication, it must also have any other root certificates necessary to validate client certificates presented on its key ring.

Application model considerations

AT-TLS support provides for several typical socket application models. Socket applications with significantly different models might not benefit from AT-TLS.

Client application model

As shown in Figure 140 on page 1141, this type of application runs entirely within the security environment of a single user. Many users might use the application, but each usage is independent, runs in a separate process, and should not share System SSL information with other processes. All socket calls for each usage of the application are made from the same z/OS UNIX process. Most connections are active connections initiated with the connect() service by this application to a server. Some client applications, such as web browsers, repeatedly connect to servers at the same or different IP addresses.

Some client applications, such as FTP or REXEC, support a second parallel connection with the server. This is often a passive connection, established by binding to an ephemeral port and listening for a single connection back from the server's IP address. For a description of an alternative method of mapping policy for these special cases, see "Secondary connection application model" on page 1145.

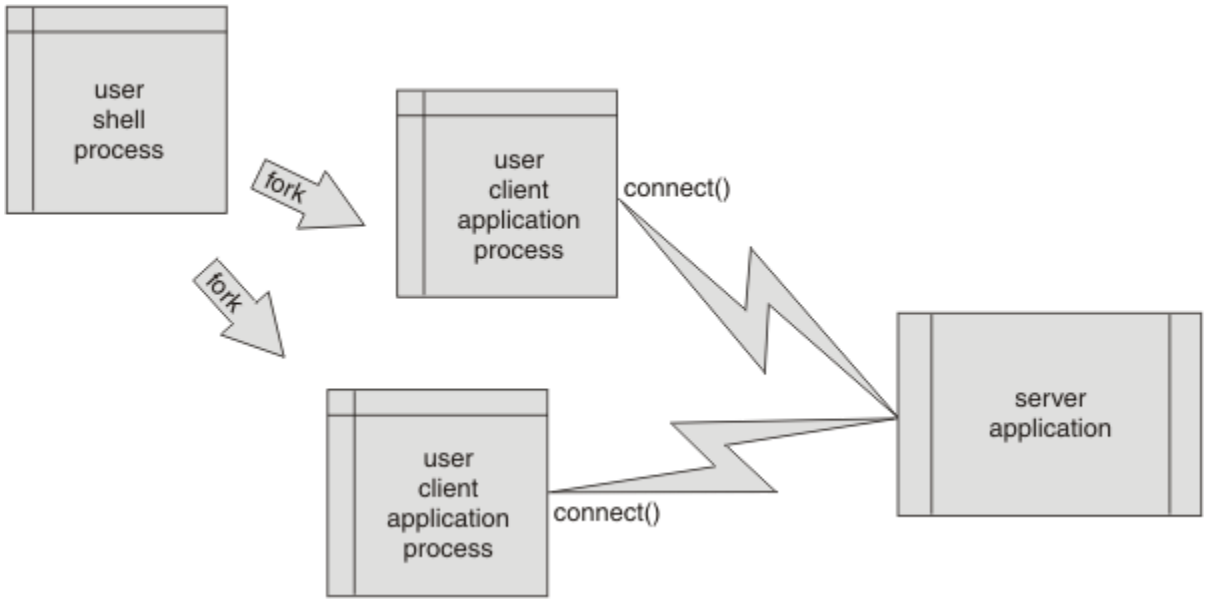


Figure 140. Client application model

All connections established by this application, whether active or passive, perform TLS handshake processing as a client. All of the connections established by a single process under the same user ID should be able to share the session ID or session ticket cache associated with the TTLSEnvironment.

If the server does not require client authentication, this client can use a virtual CERTAUTH keyring (*AUTH*/*) or shared key ring containing only the root certificates needed to validate the server's certificate.

If the server does require client authentication, each user ID needs to own a key ring that contains:

- the root certificate needed to validate the server's certificate
- the user's client certificate and private key
- the certificate authority certificates for the client's own trust chain

Tip: SAF key ring names are qualified with the owning user ID and are specified as *userid/keyring*. If the user ID is not explicitly specified on the AT-TLS key ring parameter, when it is time to open the keyring, System SSL will look for a key ring with the specified name that is owned by the user ID under which the client is executing. Therefore, specifying a key ring name without the owning user ID on the AT-TLS key ring parameter allows you to define a single TTLSEnvironmentAction keyring parameter that represent all clients. As long as each of those client user IDs own a key ring of the specified name, System SSL will use the client's own key ring.

Server application model

As shown in Figure 141 on page 1142, this type of application runs entirely within a single z/OS UNIX process. Connections can be either passive connections returned from a listening socket by the accept() service, or active connections initiated with the connect() service. Communication partners can be client applications or peer servers. Connections can be processed by subtasks or pthreads within the server process. The initial read or write of data on the connection is done under the primary security environment of the server process. Some server applications allow a client to log in with a user ID on the server system and can place this client-specified identity on the subtask or pthread used to access resources on behalf of the client. The user ID associated with the server is used for AT-TLS purposes, regardless of this ability to change to the client-specified identity.

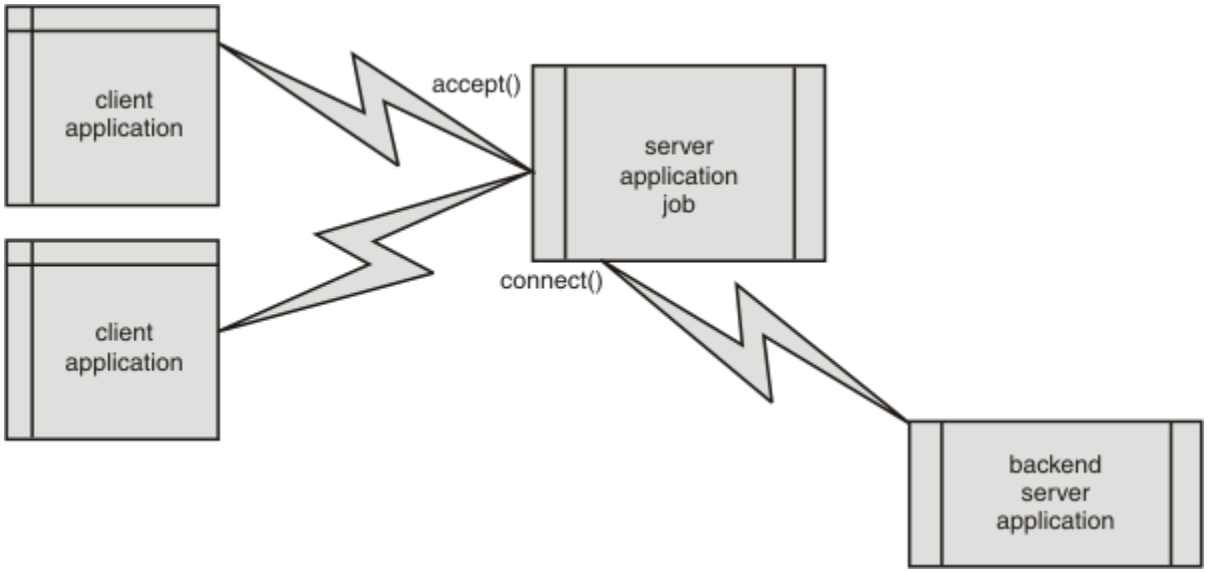


Figure 141. Server application model

Connections established by this application, whether active or passive, can perform TLS handshake processing as either a client or server. All of the connections established by a single process, under the same user ID and performing the handshake as a client or as a server, should be able to share a session ID or session ticket cache in the SSL environment. The server process should use a private key ring with the server's personal certificate and associated private keys. The key ring used must also contain the chain of signing (Certificate Authority) certificates needed to validate the server certificates it presents to its client. If the server requires client authentication, its key ring must also have any other root certificates necessary to validate the client certificates presented during TLS handshakes.

Forked server application model

As shown in Figure 142 on page 1143, this type of application is forked by a daemon application, such as INETD, to handle a single passive connection to a socket that the daemon is listening on. The daemon invokes the bind(), listen(), and accept() services on the parent socket. It then forks a new process to handle each child connection, optionally changes the new process to a different identity, and optionally execs a configured server application. The server application reads and writes data on the child connection.

Some server applications support a second parallel connection with the client. This is often an active connection, established by connecting back to an ephemeral port opened by the client at the client's IP address. For a description of an alternative method of mapping policy for these special cases, see “Secondary connection application model” on page 1145.

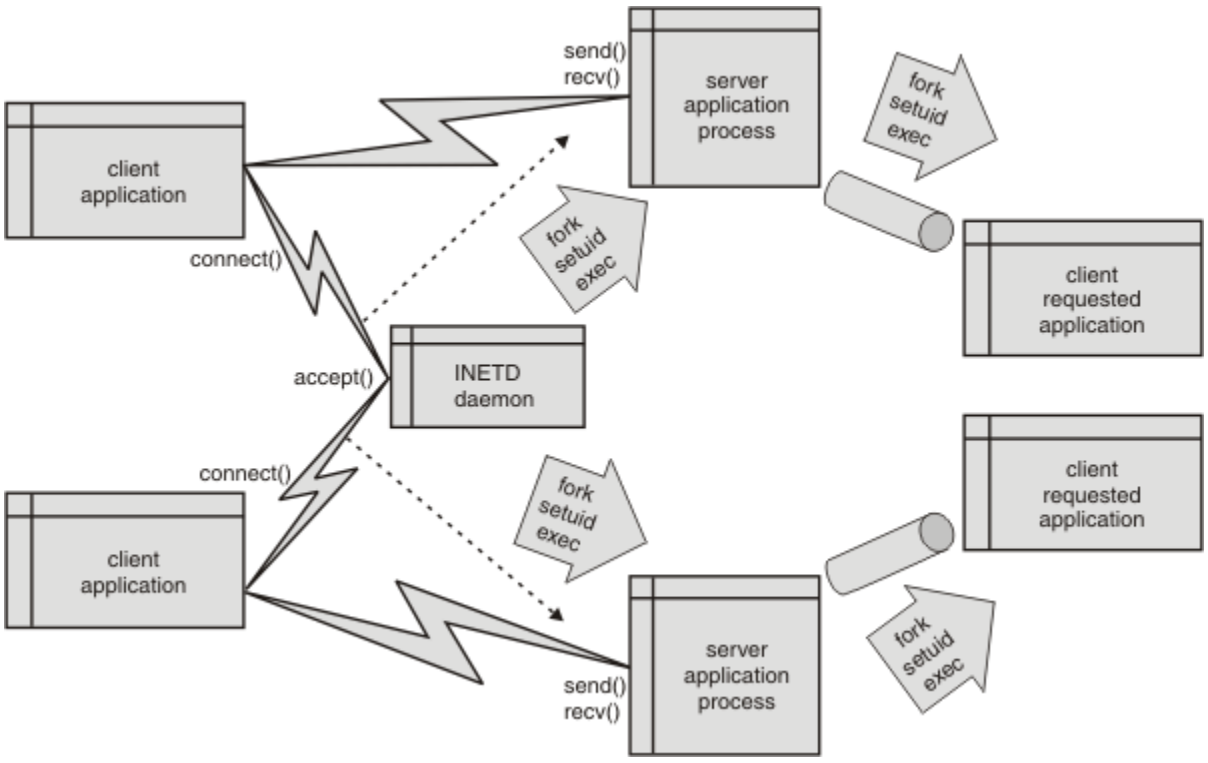


Figure 142. Forked server application model

Many server applications allow the client to log in with a user ID on the server system. Some server applications change the security environment of the server process to the client-specified identity prior to accessing resources on behalf of the client. Other server applications set up another communications path (pseudo terminal, named pipe, UNIX domain socket, and so on) and then fork an additional client process. The server changes the security environment in this client process to the client-specified identity, and then execs a client-specified program or login shell.

The initial child connection transferred by the daemon to the server process, and any additional connections established by the server process, perform TLS handshaking as the server, optionally requiring client authentication. Each forked server process runs under the same user ID. All of the server processes forked to handle child connections of the same parent connection should be able to share the session ID or session ticket cache in the SSL environment. The server processes can use a shared key ring with a site certificate and private key, or a private key ring with the server's personal certificates and associated private keys. The key ring used must also contain the chain of signing (Certificate Authority) certificates needed to validate the site or server certificates it presents to the its client. If the server requires client authentication, its key ring must also have any other root certificates necessary to validate client certificates presented during TLS handshakes.

Any client process forked by the server process is treated as a new application by AT-TLS. New connections established will not share an SSL environment with the server process that created them.

CICS transaction model

TCP/IP CICS socket support provides a CICS resource manager that invokes z/OS UNIX socket services. It also provides a CICS Listener transaction that accepts passive connections from a listening socket. Each listening socket has a configured transaction that is launched to process a single established connection.

For more information on configuring TCP/IP CICS socket support, see z/OS Communications Server: IP CICS Sockets Guide.

Advanced application considerations

Some applications need to be aware of when a secure connection is being used or examine the certificate presented by the partner. Other applications need to control if and when the TLS handshake occurs. These applications typically support both TLS and non-TLS connections over the same port. They define an application protocol for negotiating whether to use TLS and when to begin. In both cases, these applications need to be aware that TLS is being used on the connection. However, you might not want to, or might not be able to, use any SSL toolkits in the application. AT-TLS support provides the SIOCTTLSCTL ioctl commands that can be used in these situations.

Some applications establish a second connection using ephemeral ports or after the server has changed to a client supplied identity. These secondary connections need to be associated with the policy and security environment used on the primary connection. AT-TLS provides special support for these applications.

AT-TLS aware application considerations

Applications that need to examine the partner's certificate can issue the SIOCTTLSCTL IOCTL with request type TTLS_RETURN_CERTIFICATE to get the certificate at any time during a secure connection. Applications that are running under a policy with the HandshakeRole parameter set to CLIENT receive the server's certificate. Applications that are running under a policy with the HandshakeRole parameter set to ServerWithClientAuth receive the client's certificate if provided.

Applications configured as HandshakeRole ServerWithClientAuth that need to examine or use the user ID associated with the certificate in SAF can issue the SIOCTTLSCTL ioctl with request type TTLS_QUERY_ONLY or TTLS_RETURN_CERTIFICATE. If a partner certificate is available on the secure connection, AT-TLS uses a RACF service to extract the associated user ID. If no client certificate is available, or no user ID has been associated, the ioctl returns zero as the associated user ID length.

AT-TLS controlling application considerations

Applications that need to control AT-TLS behavior, using the SIOCTTLSCTL IOCTL with the TTLS_INIT_CONNECTION, TTLS_RESET_SESSION, TTLS_RESET_CIPHER, TTLS_RESET_WRITE_CIPHER, or TTLS_SEND_SESSION_TICKET request flags, must have the ApplicationControlled parameter set to ON in their TTLSEnvironmentAdvancedParms or TTLSConnectionAdvancedParms statement. This causes AT-TLS to postpone the TLS handshake. After the connection is established, the application can issue the SIOCTTLSCTL IOCTL to get the current AT-TLS connection status and determine whether or not AT-TLS support is available on this connection. When the application is ready for AT-TLS to perform the TLS handshake, it issues the SIOCTTLSCTL IOCTL with request type TTLS_INIT_CONNECTION. The SIOCTTLSCTL IOCTL initiates an AT-TLS policy lookup, if one has not yet been done, and assigns a rule and actions to the connection if a match is found. For more IOCTL information, see z/OS Communications Server: IP Programmer's Guide and Reference.

Some application protocols provide a way for the client and server programs to negotiate whether to use the TLS or SSL protocol to protect data on the connection. This cleartext negotiation typically occurs very early in the connection. When both partners agree to use TLS, they initiate the handshake. These applications can take advantage of AT-TLS. The policy must indicate ApplicationControlled ON. After the connection is established, the application can use the SIOCTTLSCTL IOCTL to determine whether AT-TLS support is configured, policy covers the connection, and the policy specifies ApplicationControlled ON. The application can then send and receive cleartext data to negotiate the use of TLS. When both partners agree, they each must initiate a secure connection. The AT-TLS application can use the SIOCTTLSCTL ioctl with TTLS_INIT_CONNECTION to initialize the secure connection. AT-TLS performs the initial handshake and provides encryption and decryption services for the application. The application can simply send and receive cleartext over the socket as it would if it were not a secure connection.

The application can also use the SIOCTTLSCTL IOCTL to reset the cipher or the session. Resetting the session clears any cached session IDs or session tickets preventing session resumption with this session’s credentials. Resetting the cipher causes the encryption key to be updated.

- For TLSv1.2 and earlier versions, it causes AT-TLS to initiate another handshake. If the session is reset first, the subsequent handshake is a full handshake. If the session has not been reset, a short handshake is attempted for the subsequent handshake. The partner application must agree to the short handshake.
- For TLSv1.3, there are two types of reset requests that can be specified on the SIOCTTLSCTL IOCTL – TTLS_RESET_CIPHER and TTLS_RESET_WRITE_CIPHER. TTLS_RESET_CIPHER causes the encryption key to be updated and a Key Update message to be sent to the session partner, requesting that the session partner also update its encryption key. The TTLS_RESET_WRITE_CIPHER causes the encryption key to be updated and a Key Update message to be sent to the session partner but there is no request that the session partner update its encryption key.

For TLSv1.3, the application can also use the SIOCTTLSCTL IOCTL to request that a session ticket be sent from the server to the client. The TTLS_SEND_SESSION_TICKET request can be used when the server is configured to enable session tickets but the count of session tickets automatically sent when the initial handshake completes is 0 (GSK_SESSION_TICKET_SERVER_ENABLE and GSK_SESSION_TICKET_SERVER_COUNT on the TTLSGskAdvancedParms statement). The session ticket is cached and used for session resumption.

Secondary connection application model

Some applications create two connections between the client and server programs. These applications typically have a single primary connection that is bound to a well-known port on the server side. After exchanging some information, a second connection is established. This second connection often uses dynamically assigned ports on both ends. Examples of this behavior include the stderr connection in the rsh, rexec, and rlogin family of applications, as well as the firewall-friendly FTP data connection. It is often not possible to define a set of policy rule conditions to correctly map these secondary connections on the client side or when the server forks a new process, with a dynamic job name, for each connected client.

In other cases, this second connection is established after the server has changed to a client-supplied identity. Mapping this second connection to AT-TLS policy as a new and independent connection would force the use of a different System SSL environment. The client-supplied identity would need to have access to the certificate private keys. Normal FTP data connections are an example of this behavior.

AT-TLS provides an alternate method of mapping policy for these secondary connections. This alternate method causes the secondary connection to share the System SSL environment and security environment of the associated primary connection.

To activate the alternate policy mapping method, define a policy rule using conditions that will map the primary connection. In this policy, specify the SecondaryMap parameter with a value of ON. When this policy is mapped to a primary connection, an entry is made in an internal table. Future connections do a normal policy lookup, and then look in the internal table for an entry with the same process ID and pair of IP addresses. If a matching entry is found and the new connection has no mapped policy, or has a mapped policy with a lower priority than the matching entry, the new connection is marked as a secondary connection and uses the same policy and user ID as the primary connection.

You should use this alternate policy mapping method only for client applications and server applications that have a single primary connection. Careful consideration should be given before using it for non-forking server applications that accept multiple primary connections, such as MVRSHD (TCP/IP's combined rsh and rexec server for the TSO environment). The alternative method of policy mapping always associates secondary connections with the most recent primary connection mapped by this process. When the process establishes multiple primary connections, the alternate mapping method is not able to reliably associate secondary connections with the correct primary connection. You should not use this alternate policy mapping method when the primary connections can map to different policies based on client IP address or multiple server listening port numbers. You should use normal policy mapping with a job name condition for the secondary connections of non-forking servers.

Chapter 21. z/OS Load Balancing Advisor

The z/OS Load Balancing Advisor communicates with external load balancers and one or more Load Balancing Agents. The main function of the Load Balancing Advisor is to provide external TCP/IP load balancing solutions, such as the Cisco Content Switching Module (CSM), with recommendations on which TCP/IP applications and target z/OS systems within a z/OS sysplex are best equipped to handle new TCP/IP workload requests. These recommendations can then be used by the load balancer to determine how to route new requests to the target applications and systems (that is, how many requests should be routed to each target). The recommendations provided by the Advisor are dynamic, and can change as the conditions of the target systems and applications change. The recommendations include several key components:

- State of the target application and system

  This includes an indication of whether the target application and target system is active. This enables the load balancer to exclude systems that are not active or do not have the wanted application running.
- z/OS Workload Management (WLM) system-wide recommendations

  WLM recommendations provide a relative measure of a target system's ability to handle new workload, as compared to other target systems in the sysplex. The WLM recommendations are derived using several measures, including each system's available general CPU capacity, which is used for both system members and application members. For application members, the amount of available System z Application Assist Processor (zAAP) capacity and System z Integrated Information Processor (zIIP) capacity can also be considered.

  If systems are 100% used, a WLM recommendation is a measurement of available displaceable capacity (capacity that can be displaced by higher importance workloads). The latter is important for scenarios where systems might be 100% used, but some might be running a larger portion of lower importance work (as defined by the WLM policy) that can therefore be displaced by higher importance workloads.
- z/OS WLM server-specific recommendations

  These recommendations are similar to the WLM system-wide recommendations, but are more specific as they are based on the following factors:

  – How individual server applications are doing compared to the WLM policy goals that are specified for that workload.
  – The amount of displaceable capacity of the general, zAAP, and zIIP processor work on each system, which is based on the following factors:

    - The workload's importance (as defined by the WLM policy).
    - The proportion of each processor type that is being consumed by the application's workload.

  These recommendations can help the load balancers avoid selecting application servers that are experiencing performance problems (that is, not meeting the specified WLM policy goals).
- Application server health from a TCP/IP perspective

  TCP/IP statistics for target applications are monitored to determine whether specific server applications are encountering problems keeping up with the current workload. For example, is a target TCP server application keeping up with TCP connection requests? Or are requests being rejected because the backlog queue is full? In scenarios where this occurs, the recommendations passed back to the load balancers are adjusted appropriately, so that the load balancer can direct fewer connections to any application that is experiencing these problems. These recommendations are provided for both UDP and TCP server applications. These recommendations are referred to as Communication Server weights in this information.

Figure 143 on page 1148 illustrates the relationship between the load balancer, a z/OS Load Balancing Advisor, and Load Balancing Agents.

© Copyright IBM Corp. 2000, 2022

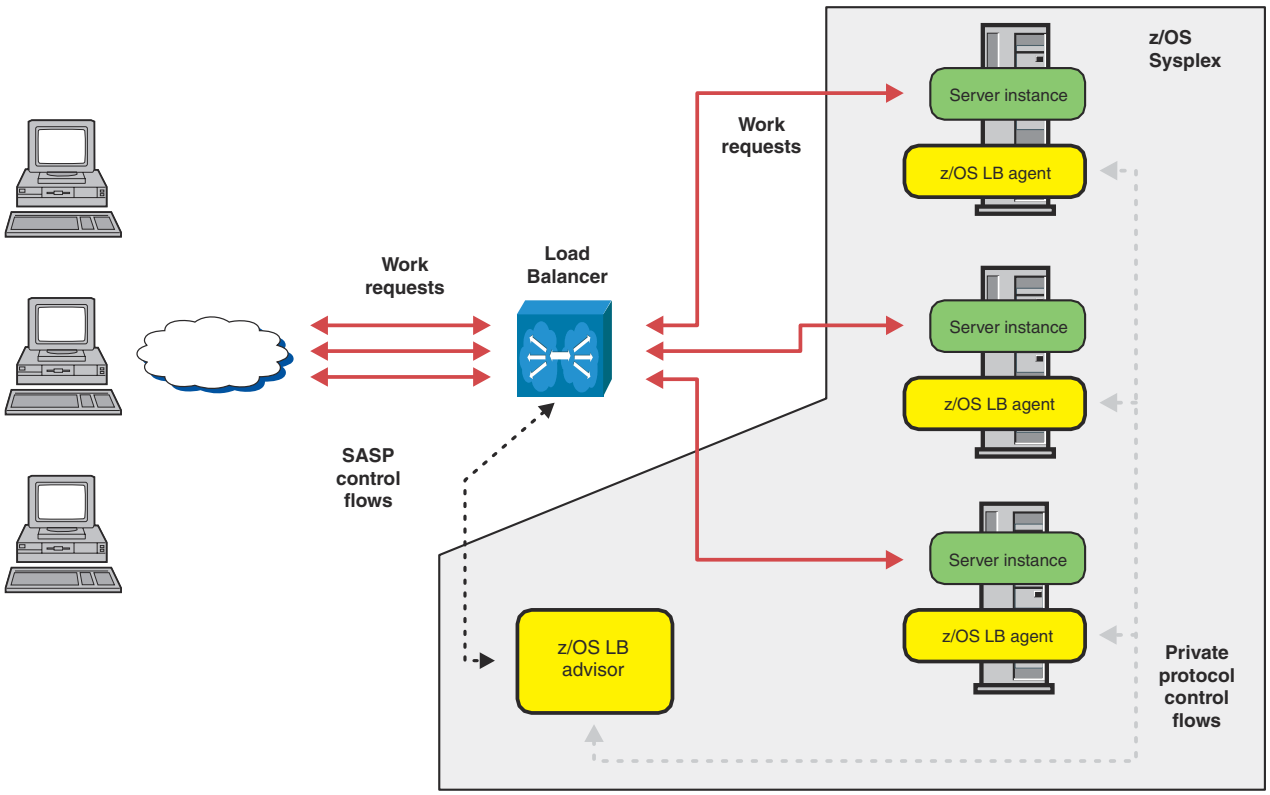


Figure 143. z/OS Load Balancing Advisor

The load balancer is configured with a list of systems and applications that it will balance. The load balancer tells the Load Balancing Advisor about the applications by specifying an IP address, port, and protocol, or about the systems by specifying an IP address. The Advisor is configured with a poll interval at which the Agents update the Advisor's data. You can configure the Advisor with a list of authorized load balancers and a list of authorized Load Balancing Agents with which it can gather data, or you can use AT-TLS support to provide the authorization for the load balancers and Load Balancing Agents. Each Agent gathers data on its own z/OS system about the TCP/IP stacks and applications running on that system.

The Agent is configured with the information it needs to contact the Advisor. The Advisor consolidates the data from all its Agents, and returns the data to the load balancer to advise the load balancer about the status of the systems and applications.

Before using the z/OS Load Balancing Advisor, you need to consider the characteristics of your environment, such as whether you need to use TLS/SSL (using AT-TLS on z/OS), what TCP/IP workloads you want to load balance, what load balancing solution you will use, and so on, as described in "Steps for preparing to use the z/OS Load Balancing Advisor" on page 1148.

Steps for preparing to use the z/OS Load Balancing Advisor

This topic describes the preparation needed to use the z/OS Load Balancing Advisor, such as selecting a load balancing solution and considering who has authority to start the Advisor and Agents.

Before you begin

You must meet the following requirements:

- You must have at least one external load balancer that supports the Server/Application State Protocol (SASP). This load balancer must have IP connectivity to each z/OS system in the sysplex that is to participate in load balancing. If you are using TLS/SSL (through AT-TLS on z/OS) for incoming connections to the Load Balancing Advisor, the load balancer must also be capable of using TLS/SSL on its SASP communication flows.

- Read Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121 and Chapter 14, “Policy-based networking,” on page 787 to decide how you want the Advisor and Agents to use AT-TLS policies. To use AT-TLS, AT-TLS must be enabled and the Policy Agent must be configured and activated for each TCP/IP stack where the Load Balancing Advisors and Load Balancing Agents might run.

Procedure

Perform the following steps to prepare to use the z/OS Load Balancing Advisor:

1. Consider whether you need to use TLS/SSL (using AT-TLS on z/OS).
   See “Step 1: Consider whether to use TLS/SSL (using AT-TLS on z/OS)” on page 1149.
2. Evaluate TCP/IP workloads to be load balanced and select a load balancing solution. (optional)
   See “Step 2: Evaluate TCP/IP workloads to be load balanced and select a load balancing solution (optional)” on page 1150.
3. Decide who will have authority to start the Advisor. (optional)
   See “Step 3: Decide who will have authority to start the Advisor (optional)” on page 1150.
4. Decide who will have authority to start the Agents. (optional)
   See “Step 4: Decide who will have authority to start the Agents (optional)” on page 1151.
5. Authorize the Agents to use WLM services.
   See “Step 5: Authorize the Agents to use WLM services” on page 1151.
6. Determine how the Advisor and agent are to interact in a subplexing environment (optional)
   See “Step 6: Determine how the Advisor and agent are to interact in a subplexing environment (optional)” on page 1152.

Step 1: Consider whether to use TLS/SSL (using AT-TLS on z/OS)

As you plan to use the z/OS Load Balancing Advisor, consider whether you need to use TLS/SSL (using AT-TLS on z/OS). The z/OS Load Balancing Advisor acts as a TCP server application, listening on two distinct ports that allow both Load Balancing Agents and external load balancers [or automated domain name registration (ADNR)] to connect to it. You need to restrict the ability to establish a connection to either of these ports, because sensitive interfaces can be exploited after a connection is accepted by the Load Balancing Advisor. For the agent listening port, you need to ensure that only authorized agents are allowed to connect, because these agents are responsible for providing sensitive information that indicates server application availability, health, and performance. For the external load balancer Server/Application State Protocol (SASP) port, you need to ensure that only authorized load balancers and ADNR are allowed to connect, because this interface can be used to obtain sensitive information regarding TCP/IP applications in a sysplex, CPU usage information for each system, and so on. You can use AT-TLS to encrypt data between the external load balancer and the Advisor's TCP/IP stack, and between the Agent's TCP/IP stack and the Advisor's TCP/IP stack.

You can use one or both of the following methods to authorize connections to the z/OS Load Balancing Advisor:

- You can explicitly configure the following lists:
  - The list of IP addresses of all the external load balancers (including ADNR) that are allowed to connect to the Load Balancing Advisor
  - The list of source IP addresses and source ports that each of the Load Balancing Agents use to connect to the Load Balancing Advisor
- You can establish policies using the z/OS Policy Agent so that the Agents, ADNR, or both are required to use TLS/SSL through AT-TLS, and load balancers are required to use TLS/SSL.

Although the configuration parameters might be sufficient in certain environments in which the Load Balancing Advisor, Agents, and external load balancers all are inside a secure network (that is, isolated by a firewall and so on), they might not be sufficient in environments in which the network is not considered to be as secure or in which the need to protect against IP address spoofing attacks is important.

With AT-TLS, the z/OS Load Balancing Advisor provides you with a more secure way to authorize access to critical Load Balancing Advisor resources using industry-standard network security standards like TLS/SSL. The AT-TLS approach also provides some additional benefits:

- Ease of use
  - Reduces the number of IP address and port lists that need to be maintained in the Load Balancing Advisor and coordinated with the external load balancers and Load Balancing Agents.
  - Eliminates the need for defining a source IP address and port for each Load Balancing Agent. This includes the Agent configuration file, the TCP ports that need to be reserved in the TCP/IP profile, and the DVIPAs that are recommended for the source IP address that is used for Agent connections.
- Outage avoidance

  If you do not use AT-TLS, adding an Agent instance into the sysplex requires updates to the Load Balancing Advisor configuration, which in turn requires a recycle of the Load Balancing Advisor so that it can process the configuration changes. With AT-TLS, you can add an Agent instance into the sysplex without recycling the Advisor.

For more information about using AT-TLS, see Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121.

Step 2: Evaluate TCP/IP workloads to be load balanced and select a load balancing solution (optional)

The first steps involve identifying the TCP/IP applications that you want to load balance, the systems that these applications will be running on, and ensuring that these applications can exploit load balancing. When this is done, evaluate the load balancing techniques that best meet your requirements. You might want to use a combination of workload balancing solutions.

There are various technology choices for performing IP load balancing in a sysplex environment. For a discussion of some of these choices, see "Workload balancing" on page 468.

Step 3: Decide who will have authority to start the Advisor (optional)

Explicit authority should be granted to all users that can start the Advisor, to prevent unauthorized users from starting it. If you do not grant explicit authority, any user able to issue the START command can start the Advisor.

Tip: You might want to combine this step with the next two steps, "Step 4: Decide who will have authority to start the Agents (optional)" on page 1151, and "Step 5: Authorize the Agents to use WLM services" on page 1151, because these steps use similar commands.

Steps for granting authority to start the Advisor

Follow these steps after you decide who you want to be able to start the Advisor.

Procedure

Perform the following steps to grant authority to start the Advisor:

1. Ensure that the OPERCMDS class is active and RACLISTed, and RACLIST processing is enabled:

   ```
   SETROPTS CLASSACT(OPERCMDS)
   SETROPTS RACLIST (OPERCMDS)
   ```

2. Define the following OPERCMDS class profile using a security product like RACF:

   ```
   RDEFINE  OPERCMDS (MVS.SERVMGR.LBADV) UACC(NONE)
   ```

3. Grant the Advisor access to the OPERCMDS class:

```
PERMIT   MVS.SERVMGR.LBADV CLASS(OPERCMDS) ACCESS(CONTROL) -
                ID(userid)
```

4. Refresh the OPERCMDS class:

```
SETROPTS RACLIST(OPERCMDS) REFRESH
```

5. See the EZARACF sample in SEZAINST for specific instructions.

   All commands that you can issue against the Advisor are MODIFY commands, with the exception of the STOP command used to stop the Advisor. Therefore, you might also want to limit which users are able to issue MODIFY and STOP commands.

Step 4: Decide who will have authority to start the Agents (optional)

Explicit authority should be granted to all users that can start the Agents, to prevent unauthorized users from starting them. If you do not grant explicit authority, any user able to issue the START command can start the Agents.

Steps for granting authority to start the Agents

Follow these steps after you decide who you want to be able to start the Agents.

Procedure

Perform the following steps to grant authority to start the Agents:

1. Ensure that the OPERCMDS class is active and RACLISTed, and RACLIST processing is enabled.

   If you have already done this for the Advisor, you can skip this step.

```
SETROPTS CLASSACT(OPERCMDS)
SETROPTS RACLIST (OPERCMDS)
```

2. Define the following OPERCMDS class profile using a security product like RACF:

```
RDEFINE  OPERCMDS (MVS.SERVMGR.LBAGENT) UACC(NONE)
```

3. Grant the Agents access to the OPERCMDS class:

```
PERMIT   MVS.SERVMGR.LBAGENT CLASS(OPERCMDS) ACCESS(CONTROL) -
                ID(userid)
```

4. Refresh the OPERCMDS class:

```
SETROPTS RACLIST(OPERCMDS) REFRESH
```

5. See the EZARACF sample in SEZAINST for specific instructions.

   All commands that you can issue against the Agents are MODIFY commands, with the exception of the STOP command used to stop the Agents. Therefore, you might also want to limit which users are able to issue MODIFY and STOP commands.

Step 5: Authorize the Agents to use WLM services

You might want or need to define the BPX.WLMSERVER resource profile to your security product and grant the Agents access to it. If you are using RACF and already have the resource profile defined and the FACILITY class is enabled, permit the Agents to the resource profile. If you are using a security product other than RACF that by default denies access to the resource profile, grant the Agents access to the resource profile. If you do not already have the resource profile defined and you are using RACF, consult the documentation of other programs and products that require WLM services and coordinate any potential changes with these programs and products.

Steps for defining the resource profile with RACF

You can define the BPX.WLMSERVER resource profile to RACF, and grant the Agents access to it.

Procedure

Perform the following steps for RACF if you choose to define the resource profile:

1. Ensure that the FACILITY class is active and RACLISTed, and RACLIST processing is enabled:

```
SETROPTS CLASSACT(FACILITY)
SETROPTS RACLIST (FACILITY)
```

2. Define the following FACILITY class profile:

```
RDEFINE  FACILITY (BPX.WLMSERVER) UACC(NONE)
```

3. Grant the Agent access to the FACILITY class:

```
PERMIT   BPX.WLMSERVER CLASS(FACILITY) ACCESS(READ) -
                       ID(userid)
```

4. Refresh the FACILITY class:

```
SETROPTS RACLIST(FACILITY) REFRESH
```

Results

For more information, see the EZARACF sample in SEZAINST.

Step 6: Determine how the Advisor and agent are to interact in a subplexing environment (optional)

If you are configuring the z/OS Load Balancing Advisor in a subplexing environment, you need to decide how you want the Advisor and Agents to interact:

- You need to determine what set of subplexes will exist in your sysplex, both how many VTAM subplexes and how many TCP/IP subplexes within a VTAM subplex.
- You need to decide which subplexes will need Load Balancing Agents and a Load Balancing Advisor.

For information about subplexing environments, see “Sysplex subplexing” on page 436.

In addition, if you are also using Automated Domain Name Registration (ADNR), see Chapter 22, “Automated domain name registration,” on page 1195 for some additional considerations for configuring ADNR in a subplexing environment.

Steps for configuring the z/OS Load Balancing Advisor

The z/OS Load Balancing Advisor provides external TCP/IP load balancing solutions with recommendations on which TCP/IP applications and target z/OS systems within a z/OS sysplex are best equipped to handle new TCP/IP workload requests.

Before you begin

You need to review “Steps for preparing to use the z/OS Load Balancing Advisor” on page 1148.

Procedure

Perform the following steps to configure the z/OS Load Balancing Advisor and one or more Load Balancing Agents.

1. Configure the Advisor and Agents to automatically restart in case of application or system failure. (optional)

See “Step 1: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)” on page 1153.

2. Configure and start syslogd.

   See “Step 2: Configure and start syslogd” on page 1155.

3. Configure one Advisor per sysplex.

   See “Step 3: Configure one Advisor per sysplex” on page 1156.

4. Configure one Agent per z/OS system in the sysplex.

   See “Step 4: Configure one Agent per z/OS system in the sysplex” on page 1161.

5. Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run. (optional)

   See “Step 5: Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run (optional)” on page 1163.

6. Customize WLM policies for the Advisor and Agents. (optional)

   See “Step 6: Customize WLM policies for the Advisor and Agents (optional)” on page 1168.

7. Configure the external load balancers.

   See “Step 7: Configure the external load balancers” on page 1168.

Step 1: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)

Although this step is optional, performing it will provide high availability to your target applications. In the event that an Agent fails, the Advisor would indicate that it has no information for any applications running on that system. As a result, target applications on the failing system would cease to receive new workload requests, in most cases, until the Agent is restarted. Automatically restarting the Agent on the same system would minimize this perceived outage. This can be accomplished using automation software or by defining an automatic restart manager (ARM) policy. For more information on defining ARM policies, see z/OS MVS Setting Up a Sysplex. In a sysplex subplexing environment, this step requires additional actions. For information about the changes to this step, see “Considerations for automatic restart in a subplexing environment” on page 1155.

The Agent registers with ARM using the following values:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=EZB<i>sysclone</i>LBAGENT
TERMTYPE=ELEMTERM
```

where *sysclone* is a 1- or 2-character shorthand notation for the name of the MVS system. For example, if the *sysclone* value is 02, the resulting ELEMNAME value is EZB02LBAGENT. For a complete description of the SYSCLONE static system symbol, see z/OS MVS Initialization and Tuning Reference.

This indicates that if the Agent fails on this system, it should be restarted on this system only.

If the Advisor or its underlying system were to fail, the load balancer might continue to distribute workload requests according to the last set of information received from the Advisor, it might resort to preconfigured weights, or it might even stop distributing new work requests to the cluster. (The behavior depends upon the load balancer implementation; consult the load balancer documentation for details.) Therefore, it is important that the Advisor be restarted as soon as possible when a failure occurs, so that it can begin communicating with the load balancer and workload request distribution can resume normally. This restart capability should cover scenarios where the Advisor itself fails, and where the system that the Advisor is running on fails. The Advisor can run on any system in the sysplex and thus can be restarted on any system in the sysplex, as long as it is configured to use dynamic VIPAs and dynamic routing is in effect. The Advisor registers with ARM using the following values:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=EZBLBADV
TERMTYPE=ALLTERM
```

This indicates that the Advisor should be restarted only on the same system in cases where the Advisor itself fails, and also restarted on a different system if the system fails. Using an ARM policy, you can indicate which systems are eligible for running the Advisor in the case of system failures. You also need to ensure that the specified backup systems have all the necessary configuration in place to enable the Advisor to be restarted there.

Some special considerations exist for scenarios where ARM is used and the TCP/IP stack address space terminates, as the result of a failure or of a planned operation. When the TCP/IP stack becomes unavailable, the Advisor also terminates, as it can no longer establish any TCP/IP communications. An ARM restart of the Advisor will likely fail, as the TCP/IP protocol stack will not be available when the restarts occur. You can handle these scenarios in the following ways:

- Planned outages of the TCP/IP stack

  Manually start the Advisor on another system, as soon as the Advisor terminates on the system where TCP/IP is stopped.
- Unplanned outages of the TCP/IP stack

  Ensure that an ARM policy (or other automation) is in place to quickly restart the TCP/IP stack on the same system. The Advisor also needs to be quickly restarted on the same system. This can be done by using an automation software package, or by using the TCP/IP profile AUTOLOG statement.

The AUTOLOG statement also has some important considerations:

- You should place the Advisor in the AUTOLOG statement list to ensure that it is started when TCP/IP is started on that system. However, you should specify the NOAUTOLOG parameter on the PORT reservation statements for the Advisor ports in the TCP/IP profile. This prevents TCP/IP from monitoring and attempting to restart the Advisor, as that could interfere with your automation logic or the ARM policy that you have put in place.
- The AUTOLOG function works best on systems where a single TCP/IP stack is active (INET environment). For CINET considerations, see "Considerations for automatic restart in a CINET environment" on page 1154.

Guideline: Establish an ARM policy with TCP/IP at a lower level than the Advisor and Agent, so that TCP/IP is restarted before the Advisor and Agent are restarted. For more information, see z/OS MVS Setting Up a Sysplex.

Requirement: The Load Balancing Advisor and Agent do not run using a system key. Therefore, if you are using ARM registration, the started task IDs need to be permitted with UPDATE authority to the associated IXCARM.SYSTCPIP.EZBLBADV and IXCARM.SYSTCPIP.EZBLBAGENT profiles in the FACILITY class within the SAF product on your system. To enable the Advisor and Agent to register with ARM, use the following RACF commands to define the profiles and grant update access to the user IDs that are assigned to start the Advisor and Agent:

```
RDEFINE FACILITY IXCARM.SYSTCPIP.EZBLBADV UACC(NONE)
RDEFINE FACILITY IXCARM.SYSTCPIP.EZBLBAGENT UACC(NONE)
PERMIT IXCARM.SYSTCPIP.EZBLBADV CLASS(FACILITY) ID(LBADV) ACCESS(UPDATE)
PERMIT IXCARM.SYSTCPIP.EZBLBAGENT CLASS(FACILITY) ID(LBAGENT) ACCESS(UPDATE)
SETROPTS RACLIST(FACILITY) REFRESH
```

Restrictions:

- If using AUTOLOG for the Agent, code the NOAUTOLOG parameter on the PORT reservation statement for the Agent port in the TCP/IP profile. This prevents the Agent from automatically being cancelled and restarted because the Agent does not listen on the port.
- If the Advisor is using IPv6 for the load balancer connections, or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to IPv6–enabled TCP/IP stacks.

Considerations for automatic restart in a CINET environment

If you are considering using the AUTOLOG statement to restart the Advisor in a CINET environment, and you placed the Advisor in the AUTOLOG statement list of each TCP/IP stack, each stack attempts to start

the Advisor during initialization. Only the first one will succeed, as only a single instance can be active at any time within a system or within a sysplex.

Considerations for automatic restart in a subplexing environment

When running the Load Balancing Advisor in a subplexing environment, there is one Load Balancing Advisor for each subplex that participates in load balancing. The subplex is determined by both the VTAM and TCP/IP subplex group IDs, which are denoted by *vv* for VTAM and *tt* for TCP/IP. These subplex group IDs are reflected in the TCP/IP sysplex group name, which is of the format EZBT*vvtt*. Each Load Balancing Advisor registers with ARM with the following parameters:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=EZBTvvttLBADV
TERMTYPE=ALLTERM
```

You must define an ARM policy. For the EZBT*vvtt*LBADV element name, you must specify the TARGET_SYSTEM keyword to indicate the systems on which the Advisor can be restarted. This ensures that the Load Balancing Advisor for a subplex is restarted only on a system that is in the same subplex. That is, it is restarted on a system that has a VTAM that was started with the same XCFGRPID (*vv*) and that has an available TCP/IP stack with the same XCFGRPID (*tt*).

In a subplexing environment, there must be one Load Balancing Agent per subplex that participates in load balancing on each z/OS system. Each Load Balancing Agent registers with ARM with the following parameters:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=EZBsysclonevvttLBAGENT
TERMTYPE=ELEMTERM
```

where:

- *sysclone* is a 1- or 2-character shorthand notation for the name of the MVS system. For a complete description of the SYSCLONE static system symbol, see z/OS MVS Initialization and Tuning Reference.
- *vvtt* is the last 4 characters of the sysplex_group_name parameter in the Agent configuration file. If this parameter is not specified, *vvtt* is omitted.

For example, if the *sysclone* value is 02 and the sysplex_group_name is EZBTCPCS, the resulting ELEMNAME value is EZB02CPCSLBAGENT.

Requirement: When ARM registration is used, the started task IDs for each Agent and each Advisor must be permitted with UPDATE authority to the IXCARM.SYSTCPIP.*elemname* profiles in the FACILITY class in the SAF-compliant security product on your system. The *elemname* value is the EZBT*vvtt*LBADV value or the EZB*sysclonevvtt*LBAGENT value previously described. You can use the following RACF commands to define the profiles and grant update access to the user IDs that are assigned to the Advisors and Agents. For each Advisor:

```
RDEFINE FACILITY IXCARM.SYSTCPIP.EZBTvvttLBADV UACC(NONE)
PERMIT IXCARM.SYSTCPIP.EZBTvvttLBADV CLASS(FACILITY) ID(advisor_userid) ACCESS(UPDATE)
SETROPTS CLASSACT(FACILITY)
```

For each Agent:

```
RDEFINE FACILITY IXCARM.SYSTCPIP.EZBsysclonevvttLBAGENT UACC(NONE)
PERMIT IXCARM.SYSTCPIP.EZBsysclonevvttLBAGENT CLASS(FACILITY) ID(agent_userid) ACCESS(UPDATE)
SETROPTS CLASSACT(FACILITY)
```

Step 2: Configure and start syslogd

The Advisor and Agent write most log messages and trace data to the syslog daemon (syslogd). A limited number of messages are written to the MVS console, but these are unaffected by syslogd configuration. For the Advisor and Agent to be able to write their log messages and trace data to syslogd, syslogd must be properly configured and started before the Advisor and Agent are started.

Because it is likely that you will be running an Agent on the same system as the Advisor, for better readability, you might want to configure syslogd to place Advisor and Agent log output in separate files. For further information, see "Configuring the syslog daemon" on page 213. In a sysplex subplexing environment, this step requires additional actions. For information about the changes to this step, see "Syslogd considerations in a subplexing environment" on page 1156.

Tip: As more data is logged by the Advisor and Agent, performance of the Advisor and Agent can be adversely affected. The amount of data that is logged by the Advisor and Agent is determined by the debug_level statement. Backing the syslogd output file with a z/OS File System instead of an HFS file system can minimize performance impacts caused by logging.

Syslogd considerations in a subplexing environment

When you are using subplexing, if you will be starting more than one instance of an Advisor or Agent on the same system, you can configure the syslog daemon (syslogd) to place the output from the different instances of the Advisor and Agent into separate files based on job name, or you can use the syslogd `-u` start option to cause the user ID and job name to be displayed on each line of the syslog. For more information, see "Configuring the syslog daemon" on page 213.

Step 3: Configure one Advisor per sysplex

There can be only one Advisor active in the sysplex at any given time, unless you are using sysplex subplexing. In a sysplex subplexing environment, this step requires additional actions. For information about the changes to this step, see "Configuring Advisors in a subplexing environment" on page 1160.

The Advisor reads configuration data from one file, which might exist as a z/OS UNIX file, a PDS or PDSE member, or a sequential data set. If you plan on allowing the Advisor to move within the sysplex in the case of failure, you probably want the Advisor configuration file or data set to exist on shared DASD, to make it accessible to all systems in the sysplex if necessary. The Advisor configuration file is specified on the CONFIG DD statement in the Advisor start procedure. A sample start procedure is provided in EZBLBADV in SEZAINST.

The Advisor configuration file serves three basic purposes:

- Defines the listening sockets for the load balancers and Agents
- Provides an access control list for specifying which load balancers and Agents can connect to the Advisor
- Customizes some optional parameters

A sample Advisor configuration file is provided in EZBLBADC in SEZAINST.

Define listening sockets/ports (required)

The Advisor maintains at least two, and up to three, listening sockets/ports, one for Agents to connect to and up to two for load balancers to connect to. There are separate IPv4 and IPv6 listening sockets for load balancers. If your TCP/IP stack is not IPv6 enabled, you will not be able to use the IPv6 listening socket.

The Advisor and Agent statements define addresses and ports on the local system and on remote systems. At times, it can be difficult to remember which statements refer to local sockets and which statements refer to remote addresses and ports.

Tip: Any statement containing the word `connection` refers to a local socket, and any statement containing the word `id` refers to a remote address and possibly a port.

Specify the local IPv4 address and port that the Advisor listens on for IPv4 load balancer connections with the lb_connection_v4 configuration statement. The default port for communications with load balancers is 3860.

The lb_connection_v6 statement does the equivalent for IPv6 that lb_connection_v4 does for IPv4. You can specify either or both of these statements. For CINET considerations regarding stack termination, see "Configuring one Advisor per sysplex in a CINET environment" on page 1160.

Guideline: To enable movement of the Advisor to another system in the sysplex or to another TCP/IP stack on the same system in the event of failure of the Advisor or its underlying system, use a dynamic VIPA (DVIPA) for the address specified on the lb_connection_v4 and lb_connection_v6 statements. Furthermore, make this DVIPA a unique application-instance DVIPA (defined through VIPARANGE) rather than a multiple application-instance DVIPA (defined through VIPADEFINE), to enable movement of the Advisor if the Advisor itself failed. For CINET considerations with DVIPAs, see "Configuring one Advisor per sysplex in a CINET environment" on page 1160.

Restriction: If the Advisor is using IPv6 for the load balancer connections, or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to IPv6–enabled TCP/IP stacks.

Specify the local port that the Advisor listens on for Agent connections with the agent_connection_port statement. If the Advisor's TCP/IP stack is IPv6 enabled, the Advisor opens a listening socket for Agents on the IPv6 unspecified address (::) on the port specified by this statement. This enables Agents to connect to the Advisor using either IPv4 or IPv6, and by using any address on the Advisor's system. If the TCP/IP stack is not IPv6 enabled, the Advisor opens a listening socket on the IPv4 unspecified address, 0.0.0.0. This enables Agents to connect to the Advisor using any IPv4 address on the Advisor's system.

Guideline: The port number used on the agent_connection_port statement should not be the same as the port used on the lb_connection_v4 statement or the lb_connection_v6 statement.

Rules:

- The port number specified on the agent_connection_port statement must match the port number specified on the Agents' advisor_id statement.
- If at least one IPv4 address is specified in the lb_id_list statement, the lb_connection_v4 statement must be specified. Similarly, if at least one IPv6 address is specified in the lb_id_list statement, the lb_connection_v6 statement must be specified.

Define the access control list

You can use one or both of the following methods for z/OS Load Balancing Advisor security:

- Access control list configuration statements

  The Advisor can control which load balancers and which Agents are allowed to connect to it by maintaining an access control list. The access control list specifies the remote IP address of the connecting load balancers and the remote IP address and port of the Agents that are allowed to connect.

  Specify the list of load balancers that are allowed to connect to the Advisor in the lb_id_list statement. Specify the list of Agents that are allowed to connect in the agent_id_list statement.

  Rules:

  – Specify only complete IP addresses in access control lists; subnetworks, IP prefixes, or other types of wildcards are not allowed.

  – The addresses in the agent_id_list statement must match the addresses in the host_connection statement of the Agents. For the purposes of high availability, the addresses specified in the agent_id_list statement of the Advisor and the host_connection statement of the Agents should be static or dynamic VIPAs, to tolerate individual link outages on the hosts.

  Restriction: There is a maximum limit of 100 load balancers that can be connected to an Advisor at any given time.

- Policies

  You can establish policies using the z/OS Policy Agent so that the Agents, ADNR, or both are required to use TLS/SSL through AT-TLS for connections to the Advisor, and load balancers are required to use TLS/SSL.

  When you are using AT-TLS for all connections to the Advisor, the Advisor's lb_id_list and agent_id_list statements and the Agents' host_connection statements are optional. If you use these statements, the rules for access control list configuration statements still apply. AT-TLS is an alternative to using these

statements, but you can still specify the statements. If you specify these statements and you are using AT-TLS, the statements are not required to match on the Advisor. For example, if an Agent connects using AT-TLS, the Advisor allows the connection to succeed even if the agent_id_list statement does not list that Agent.

Customizing optional statements

Customize the optional statements in the Advisor's configuration file.

The update_interval statement controls how often each Agent updates the Advisor with data, and depending upon load balancer implementation and configuration, can control how often load balancers are updated with information from the Advisor. The default value is 60 seconds. At each update interval, each Agent refreshes the Advisor with the status of each registered member for which the Agent is responsible. This information includes the status of the target application (active or not), whether it is an application member, the operator quiesce state of the member, and various weights that measure the target system and application's ability to handle additional workload requests. The lower the update interval, the more up-to-date the load balancer's data will be with respect to the target's availability and capability to handle additional workload requests. Of course, the lower the update interval, the higher the network traffic and CPU overhead is.

Depending upon load balancer implementation and configuration, the update_interval statement might also determine how often the load balancer is updated with data from the Advisor. If the load balancer supports the SASP push flag, and it has been specified in the load balancer, the Advisor sends data to the load balancer at least every update interval. Regardless of what the update interval is set to, if this push flag is supported and configured in the load balancer, certain events might cause the Advisor to update the load balancer with information before the update interval timer expires. These events include the starting or stopping of a target application, or the addition or deletion of a member's IP address on the Agent host.

Therefore, the update interval is a key factor in determining the latency period between when changes occur on the target application system, and when the load balancer is informed of them. Each Agent updates the Advisor with new information every update interval. The Advisor, in turn, updates the load balancer with changes in weights every update interval, if the load balancer supports the push flag. In addition, if the push flag is supported and configured by the load balancer, the Advisor updates the load balancer with any change in the target's availability status as soon as it discovers such a change from the Agent, instead of waiting for the update interval to expire. Therefore, when the load balancer supports and configures the push flag, the maximum amount of latency expected between a change in a member's weight and when the load balancer is informed of it is twice the update interval (that is, one update interval for the Agent to report it to the Advisor and one update interval for the Advisor to report it to the load balancer). However, on the average, it should take one update interval for a change in the target application weight to reach the load balancer.

Use the optional wlm statement to specify the default type of WLM recommendation to be used for all groups. There are two choices for this, the basewlm and serverwlm values. If you do not specify this statement, the default is basewlm. If you want a specific group of applications to use a type of WLM recommendation other than the default, you can override the default WLM recommendation type for that group on a port number basis using the port_list statement. The WLM recommendation is a key component used in determining the net weight assigned to a member.

The type of WLM recommendation represented by the basewlm value indicates the overall displaceable capacity (general, zAAP, and zIIP) of the system where the application represented by the member is, relative to the other systems in the sysplex. This is referred to as a WLM system weight recommendation. Use the optional proctype parameter with basewlm to specify the proportion of general, zAAP, and zIIP CPU that is consumed by an application's workload.

The serverwlm value represents a different type of WLM recommendation, in that it reflects how well an individual server application is performing from a WLM perspective (based on the WLM policy). This type of recommendation is a server-specific recommendation and is referred to as a server-specific WLM recommendation. Server-specific WLM recommendations are composed of two key elements:

- The amount of displaceable capacity (general, zAAP, and zIIP) available on the target system based on the importance level of the application, and the proportion of general, zAAP, and zIIP CPU that is currently being consumed by the application's workload. For example, if the application is using only general and zAAP CPU, the displaceable zIIP capacity is not considered.
- How well the application is performing compared to the WLM goals for that application workload.

In addition, WLM provides an interface that enables applications to report the following additional information:

- Abnormal transaction completion rate, or the rate of abnormal completions per 1000 total transactions
- Application health, a value in the range 0–100% (100% being optimal), representing the overall health of the application

Using this additional information, WLM might reduce the server-specific recommendation. For more information, see "Sysplex distributor" on page 475.

Evaluate whether you can use WLM server-specific distribution as an alternative to WLM system weight distribution for an application. In addition to the above reasons, server-specific distribution has the added advantage that processor proportions are automatically determined and dynamically updated by WLM, based on the actual CPU usage by the application. If you need to use system weight distribution, to determine the processor proportions to configure, study the workload usage of assist processors by analyzing SMF records, using performance monitors reports such as RMF, and so on.

System members (port and protocol are zero) always use WLM system weight recommendations and cannot be configured to consider zAAP and zIIP CPU, because the type of workload is unknown. This is true even if proctype is coded on the wlm statement.

Application members can use either type.

It is important that you choose the type of WLM recommendation that is best suited to each group of applications. Some types of applications are better suited to using WLM system weight recommendations rather than server-specific WLM recommendations. For most applications, server-specific WLM recommendations provide a more accurate way to distribute workload to their servers. However, when a server acts as an access point to applications that run in other address spaces (and therefore in a different service class), WLM system weight recommendations might be the preferred distribution method; if expected usage of general, zAAP, and zIIP processors is known, this recommendation can be further refined by using the proctype parameter. The sysplex distributor function can also use server-specific WLM recommendations or WLM system weight recommendations. For examples of some applications that might be better represented by WLM system weight recommendations, see "Sysplex distributor" on page 475.

The optional port_list statement enables you to override or specify parameters for members on a port number basis. The wlm parameter of the port_list statement enables you to override the value defined (or specified by default) on the wlm statement, for all members that use the port number specified. The actual WLM recommendation type used is still dependent upon the value specified and the z/OS level of the Agents owning the members of the group.

When selecting the type of WLM recommendation to use for a given group, it is important to consider the following requirements:

- All members in a group must specify the same type of WLM recommendation (using the wlm or port_list statement).
- To use server-specific recommendations, no Agent reporting on behalf of a member of a group can be at a release level prior to z/OS V1R7.

For any groups where these requirements are not met, the Advisor uses WLM system weight recommendations, and a warning message is written to syslogd. The main rationale behind this is that WLM system weight recommendations and server-specific WLM recommendations cannot be directly compared to one another.

The Advisor can detect dynamically whether these requirements are being met. For example, if all owning Agents of a group, except one, support server-specific WLM recommendations, and the application on that one system is brought down, the WLM recommendation type would change dynamically from

WLM system weight recommendations to server-specific WLM recommendations, provided the Advisor was configured to request server-specific WLM recommendations for that group. Similarly, if that same application is started back up, the WLM recommendation type would dynamically switch back to WLM system weight recommendations. A similar circumstance would arise if the member owned by the Agent that does not support server-specific WLM recommendations was quiesced by the z/OS operator or the load balancer administrator.

The optional debug_level statement determines how much trace data is captured in the Advisor's log file.

Restriction: In most cases, you should not customize the debug_level statement, unless directed to do so by an IBM Service representative. Adding additional types of trace data can cause the amount of data captured to become voluminous. Reducing the amount of trace data from the default might make diagnosing a problem more difficult.

For more details on the Advisor configuration file statements, see z/OS Communications Server: IP Configuration Reference.

Configuring one Advisor per sysplex in a CINET environment

You must define listening sockets and ports, as described in "Define listening sockets/ports (required)" on page 1156.

Rule: If you use a unique application-instance DVIPA for the Advisor in a CINET environment, all TCP/IP stacks on that system must code the VIPARANGE statement for that DVIPA. Alternatively, and less desirably, you can establish stack affinity to one of the TCP/IP stacks that are coded with the VIPARANGE statement for that DVIPA, if you do not have VIPARANGE coded for that DVIPA on all of the TCP/IP stacks on that system. This alternative, of course, does not enable the Advisor to be moved to another TCP/IP stack in the event of failure, unless you are able to restart the Advisor with a different start procedure that can establish stack affinity to another TCP/IP stack that has the DVIPA defined in a VIPARANGE statement. For information on the use of the _BPXK_SETIBMOPT_TRANSPORT environment variable that can be used to establish stack affinity, see "Generic server versus server with affinity for a specific transport provider" on page 46.

The lb_connection_v6 statement does for IPv6 what the lb_connection_v4 statement does for IPv4. You can specify either or both of these statements. If you run the Advisor on a CINET system, be aware that the address or addresses you choose for these statements tie the Advisor to the stack owning those addresses. Consequently, termination of that stack results in termination of the Advisor.

Configuring Advisors in a subplexing environment

When you are using subplexing, there can be more than one Advisor active in the sysplex at any given time. In fact, there should be one Advisor active for each subplex in the sysplex that you want to participate in load balancing through the Load Balancing Advisor. Each Advisor reads configuration data from a file, which can exist as a z/OS UNIX file, a PDS or PDSE member, or a sequential data set. In the configuration file for each Advisor, the sysplex_group_name statement specifies the TCP/IP sysplex group name, in the form EZBT*vvtt*, where *vv* is the VTAM subplex group ID specified on the VTAM XCFGRPID start option, and *tt* is the TCP/IP subplex group ID specified by the XCFGRPID parameter on the GLOBALCONFIG statement in the TCP/IP profile. If no VTAM subplex ID is specified when VTAM is started, then *vv* is CP. If no TCP/IP subplex ID is specified in the TCP/IP profile, then *tt* is CS. If you have a default subplex in your sysplex (that is, a subplex in which both the VTAM and TCP/IP subplex IDs are not specified), configure the Load Balancing Advisor for that subplex with a sysplex group name of EZBTCPCS.

Requirement: In a subplexing environment, the IP address of the Advisor's listening socket must exist on a TCP/IP stack belonging to the subplex that corresponds to the TCP/IP sysplex group name specified in the Advisor's configuration file. If there is more than one TCP/IP stack in a subplex, the IP address must be a DVIPA defined within a VIPARANGE statement on each of the stacks in the subplex. This enables the Advisor to connect regardless of the order that the TCP/IP stacks in the subplex are started.

Tip: In a subplexing environment, if you will have more than one Advisor started on the same z/OS system (in different subplexes), create unique start procedures for them or ensure that they have unique job names when they are started (for example, S LBADV.ADV0105 or S LBADV,JOBNAME=ADV0105).

Step 4: Configure one Agent per z/OS system in the sysplex

There can be only one Agent active per z/OS system in the sysplex at any point in time, unless sysplex subplexing is used. When operating in a sysplex subplexing environment, this step requires additional actions. For information about the changes to this step, see "Configuring Agents in a subplexing environment" on page 1163.

The Agent reads configuration data from one file, which can exist as a z/OS UNIX file, a PDS or PDSE member, or a sequential data set. The Agent configuration file is specified on the CONFIG DD statement in the Agent start procedure. A sample start procedure is provided in EZBLBAGE in SEZAINST.

The Agent configuration file serves three basic purposes:

- Defines the IP address and port that the Agent binds to for communication with the Advisor
- Identifies the location (IP address and port) of the Advisor
- Customizes optional parameters

A sample Agent configuration file is provided in EZBLBAGC in SEZAINST.

The Advisor and Agent statements define addresses and ports on the local system and on remote systems. At times, it can be difficult to remember which statements refer to local sockets and which statements refer to remote addresses and ports.

Tip: Any statement containing the word `connection` refers to a local socket, and any statement containing the word `id` refers to a remote address and possibly port.

Defining the IP address and port to bind to for communications with the Advisor

Specify the local IP address and port that the Agent binds to for communications with the Advisor on the host_connection statement. This is used as part of the Advisor's access control enforcement.

Rules:

- The IP address on the host_connection statement can be an IPv6 address, if the Agent's system is running an IPv6-enabled TCP/IP stack and the Advisor has an IPv6-enabled TCP/IP stack available.
- If an IPv4 address is specified on the host_connection statement, an IPv4 address must be specified on the advisor_id statement. Similarly, if an IPv6 address is specified on the host_connection statement, an IPv6 address must be specified on the advisor_id statement.

Guidelines:

- For simplicity and consistency, you might want to specify the same port on the host_connection statement for each Agent, and reserve the same port for the Agent on each TCP/IP stack that an Agent will run on. For more information about port reservation, see "Step 5: Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run (optional)" on page 1163.
- The address in the host_connection statement (and therefore, also in the Advisor's agent_id_list statement) should be a static VIPA. For CINET considerations regarding the host_connection statement, see "Configuring one Agent per z/OS system in the sysplex in a CINET environment" on page 1162.

The Agent host_connection statement and the Advisor agent_id_list statement are optional if AT-TLS is used for all Agent-Advisor connections. If you specify these statements, the rules and guidelines previously stated still apply. AT-TLS is an alternative to using these statements, but you can still specify the statements. If you specify these statements and you are using AT-TLS, the statements are not required to match on the Advisor. For example, if an Agent connects using AT-TLS, the Advisor allows the connection to succeed even if the agent_id_list statement does not list that Agent.

Also see "Configuring one Advisor per sysplex in a CINET environment" on page 1160 for CINET considerations regarding unique application-instance DVIPAs and stack affinity.

Restriction: If the Advisor is using IPv6 for the load balancer connections, or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to IPv6–enabled TCP/IP stacks.

Identifying the location of the Advisor (required)

Specify the location of the Advisor on the advisor_id statement. This statement contains an IP address and port that the Advisor uses to listen for connections from Agents in the sysplex.

Guideline: Use a dynamic VIPA for this address, to enable the Advisor to be moved within the sysplex in the event of a failure of the Advisor or the Advisor's underlying system.

Rules:

- The port specified on the advisor_id statement must match the port specified on the Advisor's agent_connection_port statement.
- If an IPv4 address is specified on the advisor_id statement, an IPv4 address must be specified on the host_connection statement. Similarly, if an IPv6 address is specified on the advisor_id statement, an IPv6 address must be specified on the host_connection statement.

Customizing optional statements

Similar to the Advisor, the optional debug_level statement determines how much trace data is captured in the Agent's log file.

Restriction: In most cases, you should not customize this statement, unless directed to do so by an IBM Service representative. Adding additional types of trace data can cause the amount of data captured to become voluminous. Reducing the amount of trace data from the default might make diagnosing a problem more difficult.

For more details on the Agent configuration file statements, see z/OS Communications Server: IP Configuration Reference.

Configuring one Agent per z/OS system in the sysplex in a CINET environment

If the system where the Agent is running is a CINET system, the address in the host_connection statement (and therefore, also in the Advisor's agent_id_list statement) should be a dynamic VIPA (DVIPA) to facilitate movement of the Agent to another TCP/IP stack on that system.

Rule: If you use a unique application-instance DVIPA (coded with VIPARANGE) for the Agent in a CINET environment, all TCP/IP stacks on that system must code the VIPARANGE statement for that DVIPA. When configured in this manner, if the TCP/IP stack that currently owns the DVIPA fails, the Agent remains up and automatically attempts to bind again to the DVIPA. The next available default TCP/IP stack becomes the new owner of the DVIPA. (If this is a sysplex subplexing environment, all TCP/IP stacks on that system that are within the same subplex as the Agent must code the same VIPARANGE statement for that DVIPA. This allows the Agent to reestablish connectivity with the Advisor through another stack in the same subplex on that system.) Alternatively, and less desirably (see following restriction), you can establish stack affinity to one of the TCP/IP stacks that are coded with the VIPARANGE statement for that DVIPA, if you do not have VIPARANGE coded for that DVIPA on all of the TCP/IP stacks on that system. Of course, this does not enable the Agent to be automatically moved to another TCP/IP stack in the event of failure. To recover the Agent in this type of configuration, manual intervention (or automation) is required. Because the Agent remains active if its TCP/IP stack fails, you must manually terminate the Agent. Then you must restart the Agent with a different start procedure that can establish stack affinity to another available TCP/IP stack. For information on the use of the _BPXK_SETIBMOPT_TRANSPORT environment variable that can be used to establish stack affinity, see “Generic server versus server with affinity for a specific transport provider” on page 46.

Restriction: When running in a CINET environment, establishing stack affinity between the Agent and one of the TCP/IP stacks on that system enables only resources on that TCP/IP stack to participate in workload balancing. The Agent running on that system is not aware of resources on the other TCP/IP stacks on that system. If you want to enable resources on all TCP/IP stacks in a CINET environment to participate in workload balancing, do not establish stack affinity with the Agent.

Configuring Agents in a subplexing environment

When you are using subplexing, there can be more than one Agent per z/OS system in the sysplex. In fact, there should be one Agent active for each subplex with a TCP/IP stack on a system that you want to participate in load balancing through the Load Balancing Advisor. Each Agent reads configuration data from a file, which can exist as a z/OS UNIX file, a PDS or PDSE member, or a sequential data set. In the configuration file for each Agent, the sysplex_group_name statement specifies the TCP/IP sysplex group name, in the form EZBT*vvtt*, where *vv* is the VTAM subplex group ID specified with the VTAM XCFGRPID start option, and *tt* is the TCP/IP subplex group ID specified with the XCFGRPID parameter on the GLOBALCONFIG statement in the TCP/IP profile. If no VTAM subplex ID is specified when VTAM is started, then *vv* is CP. If no TCP/IP subplex ID is specified in the TCP/IP profile, then *tt* is CS. If you have a default subplex in your system (that is, a subplex in which both the VTAM and TCP/IP subplex IDs are not specified), configure the Load Balancing Agent for that subplex with a sysplex group name of EZBTCPCS.

Tip: In a subplexing environment, if you will have more than one Agent started on the same z/OS system (in different subplexes), create unique start procedures for them or ensure that they have unique job names when they are started (for example, S LBAGENT.AGE0105 or S LBAGENT,JOBNAME=AGE0105).

Requirements:

- In a subplexing environment, the IP address used by the Agent to connect to the Advisor must exist on a TCP/IP stack belonging to the subplex that corresponds to the TCP/IP sysplex group name specified in the Agent's configuration file. If there is more than one TCP/IP stack in a subplex, the IP address must be a DVIPA defined within a VIPARANGE statement on each of the stacks in the subplex. This enables the Agent to connect regardless of the order that the TCP/IP stacks in the subplex are started.
- If you have more than one stack on a system, those stacks are not all in the same subplex, and you will be starting Load Balancing Agents on that system, the system must not be at a level prior to V1R10.

Step 5: Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run (optional)

When operating in a sysplex subplexing environment, this step requires additional actions. For information about the changes to this step, see "Customizing TCP/IP profiles in a subplexing environment" on page 1168.

There are several things to do to customize the TCP/IP profile to accommodate the Advisor and Agents.

- In the appropriate TCP/IP profiles that they will use, including the TCP/IP stacks that they could potentially move to, reserve the ports that the Advisor and Agents will use. All ports for the Advisor and Agent use the TCP protocol, and thus should be reserved for TCP. The Advisor has at least two ports, and potentially three ports, to reserve, including the ports specified on the following statements:
  - lb_connection_v4
  - lb_connection_v6
  - agent_connection_port

  For CINET considerations regarding the lb_connection_v4 and lb_connection_v6 statements, see "Customizing TCP/IP profiles in a CINET environment" on page 1168.
- If you use dynamic VIPAs for the Advisor as recommended, you need to configure the appropriate TCP/IP profiles in the sysplex for the DVIPA definition and usage. The preferred definitions would include VIPADEFINE with MOVEABLE IMMEDIATE, or VIPARANGE with MOVEABLE NONDISRUPTIVE. For more specific information, see "Using dynamic VIPAs" on page 369.

  Restriction: Do not mix sysplex distributor functions with these DVIPAs.
- If you are currently using the SHAREPORT or SHAREPORTWLM parameters on the TCP/IP profile PORT statement to enable multiple TCP applications to share the same port, some additional considerations might apply to your configuration. For example, if the TCP applications sharing the same port are also members of groups that are being reported to external load balancers with SASP, it is important to ensure that consistent criteria are used by the various load balancing components.

When using the z/OS Load Balancing Advisor, all instances of a TCP application sharing the same port on a target system are reported to external load balancers using a single member entry, and therefore, a single recommendation. This recommendation reflects the average net weight calculated for all the servers sharing the same port on a target system, and is based on the type of WLM recommendation configured on the Advisor. When the TCP connection requests reach a target TCP/IP stack and multiple applications are sharing the same port, the connections are then load balanced by TCP/IP across the multiple application server instances. How this load balancing is performed depends on whether the SHAREPORT or SHAREPORTWLM parameter is specified on the PORT statement. For more details on the PORT statement, see z/OS Communications Server: IP Configuration Reference.

Guideline: If server-specific WLM recommendations are configured within the Advisor for a given group that contains servers that share the same port on a given system or TCP/IP stack, the SHAREPORTWLM parameter should also be specified on the PORT statement in the TCP/IP profile for these servers. This enables both the external load balancers and the internal TCP/IP load balancer to operate with the same type of recommendations when load balancing work requests to these servers. Similarly, if WLM system weight recommendations were configured in the Advisor for a group, the SHAREPORT parameter would probably be more appropriate.

Enabling TLS/SSL for z/OS Load Balancing Advisor (optional)

For AT-TLS, the following customization tasks are required before starting the TCP/IP stacks and the Advisor and Agent applications:

- Enable AT-TLS in the TCP/IP stack.

  Specify the TTLS parameter on the TCPCONFIG statement in the TCP/IP profile. For additional information about AT-TLS, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121. For information about the TCPCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

- Set up authorization for the **pasearch** command, if the command will not be issued from a superuser.

  Create a SERVAUTH profile of EZB.PAGENT.*sysname*.*image*.*ptype*, where the *ptype* value is set to TTLS or to a wildcard value. For more information, see “Steps for configuring the Policy Agent” on page 808 and z/OS Security Server RACF Security Administrator's Guide.

- Enable AT-TLS configuration for the Policy Agent.

  Specify the CommonTTLSConfig and TTLSConfig statements in the Policy Agent configuration file for each stack. On the TTLSConfig statement, specify the path of the stack-specific AT-TLS policy file to be installed for the server. For additional information about the CommonTTLSConfig and TTLSConfig statements, see z/OS Communications Server: IP Configuration Reference.

- Define AT-TLS policies in new or existing Policy Agent configuration files.

  Specify the AT-TLS policies in the configuration files that are identified with the CommonTTLSConfig and TTLSConfig statements. Ensure that the Load Balancing Advisor policy definitions are defined on all systems in the sysplex on which the Advisor can run.

  The Load Balancing Advisor is a server application. For general information about setting up AT-TLS for a server, see Table 58 on page 1126.

  The following example shows the TTLSConfig policy file statements in the path file for the load balancer connections to the Advisor. Port 3860 is the default port.

```
TTLSRule                 LBAdvisorLBRule
{
   LocalPortRange            3860
   Direction                 Inbound
   TTLSGroupActionRef        LBAdvisorLBGroup
   TTLSEnvironmentActionRef  LBAdvisorLBEnvironment
}
TTLSGroupAction          LBAdvisorLBGroup
{
   TTLSEnabled               On
}
TTLSEnvironmentAction    LBAdvisorLBEnvironment
{
```

```
   TTLSKeyRingParms
   {
      Keyring                LBADV/server_key_ring
   }
   TTLSEnvironmentAdvancedParms
   {
      # TTLS will verify a user ID is associated with certificate
      ClientAuthType         SAFCheck
      ApplicationControlled  On
   }
   HandshakeRole             ServerWithClientAuth
   TTLSCipherParmsRef        RequireEncryption
   Trace                     7
}
```

In this example, all external load balancers must use TLS/SSL and supply a client certificate that will be validated in the key ring and must be associated with a user ID on the SAF-compliant security product on the local z/OS system. This type of policy allows additional finer-grain SAF checks using optional SERVAUTH profiles. You can use other, less restrictive, policies; however, if you use less restrictive policies, the Advisor, Agent, and ADNR require that you specify the configuration parameters for those connections (lb_id_list or agent_id_list statements in the Advisor configuration file, host_connection statement in the Agent configuration file, and host_connection_addr statement in the ADNR configuration file).

The following example shows the TTLSConfig policy file statements in the path file for the Agent connections to the Advisor. Port 8100 is the port that is used in the sample Advisor configuration file:

```
TTLSRule                 LBAdvisorAgentRule
{
   LocalPortRange            8100
   Direction                 Inbound
   TTLSGroupActionRef        LBAdvisorAgentGroup
   TTLSEnvironmentActionRef  LBAdvisorAgentEnvironment
}
TTLSGroupAction          LBAdvisorAgentGroup
{
   TTLSEnabled               On
}
TTLSEnvironmentAction    LBAdvisorAgentEnvironment
   {
   TTLSKeyRingParms
   {
      Keyring                LBADV/server_key_ring
   }
   TTLSEnvironmentAdvancedParms
   {
      # TTLS will verify a user ID is associated with certificate
      ClientAuthType         SAFCheck
      ApplicationControlled  On
   }
   HandshakeRole             ServerWithClientAuth
   TTLSCipherParmsRef        RequireEncryption
   Trace                     7
}

# Set of TLS Ciphers with Encryption
TTLSCipherParms RequireEncryption
{
   V3CipherSuites        TLS_RSA_WITH_RC4_128_MD5
   V3CipherSuites        TLS_RSA_WITH_RC4_128_SHA
   V3CipherSuites        TLS_RSA_WITH_DES_CBC_SHA
   V3CipherSuites        TLS_RSA_WITH_3DES_EDE_CBC_SHA
}
```

The Load Balancing Agent is a client application. For general information about setting up AT-TLS for a client, see Table 59 on page 1128.

You must configure the policy on the TCP/IP stack where the Agents will run with the same SSL protocol, key ring, and cipher suite (if encrypting data) for which the Advisor is configured.

The following example shows the TTLSConfig policy file statements for a Load Balancing Agent. On the TTLSConfig statement, specify the path of the stack-specific AT-TLS policy file to be installed for

the client. For more information about the TTLSConfig statement, see z/OS Communications Server: IP Configuration Reference. Port 8100 is the port that is used in the sample Agent configuration file.

```
TTLSRule                     LBAgentRule
{
   RemotePortRange               8100
   Direction                     Outbound
   TTLSGroupActionRef            LBAGroup
   TTLSEnvironmentActionRef      LBAgentEnvironment
}
TTLSGroupAction              LBAGroup
{
   TTLSEnabled                   On
}
TTLSEnvironmentAction        LBAgentEnvironment
{
   TTLSKeyRingParms
   {
      Keyring                  LBAGENT/client_key_ring
   }
      HandshakeRole                 CLIENT
      TTLSCipherParmsRef            RequireEncryption
      Trace                         7
}

# Set of TLS Ciphers with Encryption
TTLSCipherParms RequireEncryption
{
   V3CipherSuites        TLS_RSA_WITH_RC4_128_MD5
   V3CipherSuites        TLS_RSA_WITH_RC4_128_SHA
   V3CipherSuites        TLS_RSA_WITH_DES_CBC_SHA
   V3CipherSuites        TLS_RSA_WITH_3DES_EDE_CBC_SHA
}
```

For additional information, see:

- Chapter 14, “Policy-based networking,” on page 787.
- Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121. Table 58 on page 1126 is for the Advisor, and Table 59 on page 1128 is for the Agent and for load balancers.
- AT-TLS policy statements in z/OS Communications Server: IP Configuration Reference
- CommonTTLSConfig and TTLSConfig statements in z/OS Communications Server: IP Configuration Reference.
- Online help for the IBM Configuration Assistant for z/OS Communications Server

- Create z/OS server (Advisor) and client (Agent, ADNR, external load balancer) key rings and necessary certificate authority certificates.

  The server key ring needs to contain a server certificate, and any certificates that are used to sign it. The server needs access to the private keys of the server certificate. The client key ring needs the root certificate that is used to sign the server certificates.

  For a TLS/SSL primer and some step-by step examples, see Appendix B, “TLS/SSL security,” on page 1329. For more information about managing key rings and certificates with RACF and the RACDCERT command, see z/OS Security Server RACF Security Administrator's Guide. For detailed information about managing key rings and certificates with gskkyman, see z/OS Cryptographic Services System SSL Programming.
- Send the external load balancer's certificate to the z/OS host.
- Associate each user ID (for the Advisor, Agents, ADNRs, and external load balancers) with a certificate.

  Use the RACDCERT ADDRING command to define a key ring in RACF and to associate it with your application's user ID. Use the RACDCERT CONNECT command to connect certificates to the key ring. For detailed information about setting up your certificate environment, see z/OS Security Server RACF Security Administrator's Guide.
- Create client side certificates for the external load balancers. See the load balancer documentation for instructions.

- Define client user IDs on the TCP/IP stacks on which the Advisor will run by issuing security product commands to establish authorization for the user IDs.

  You can configure the Advisor's clients (Agents, ADNR, and external load balancers) to present security credentials, including a user ID. If you configure this, you must set up the security manager on the Advisor system to accept these credentials.

  Using a security product like RACF, perform the following steps to control access to the Load Balancing Advisor, Agents, and ADNR.

  1. Use the following commands to ensure that the SERVAUTH class is active and RACLISTed, and that RACLIST processing is enabled:

     ```
     SETROPTS CLASSACT(SERVAUTH)
     SETROPTS RACLIST (SERVAUTH)
     ```

  2. Use the following commands to define the following SERVAUTH class profiles on each system on which the Advisor might run.

     ```
     RDEFINE SERVAUTH EZB.LBA.LBACCESS.sysname.tcpsysplexgroupname UACC(NONE)
     RDEFINE SERVAUTH EZB.LBA.AGENTACCESS.sysname.tcpsysplexgroupname UACC(NONE)
     ```

     where *sysname* is the MVS system name or a wildcard (*) and *tcpsysplexgroupname* is the TCP/IP sysplex group name. If you are not using subplexing, use the default subplex identifier EZBTCPCS or a wildcard (*). For example, on system MVSSYS using the default subplex, the profile name is EZB.LBA.LBACCESS.MVSSYS.EZBTCPCS.

  3. Use the following commands to grant access to the SERVAUTH class for the user IDs associated with the Agents, ADNR, and the external load balancers on each system on which the Advisor might run:

     ```
     PERMIT   EZB.LBA.LBACCESS.sysname.tcpsysplexgroupname -
        CLASS(SERVAUTH) ACCESS(READ) ID(userid)
     PERMIT   EZB.LBA.AGENTACCESS.sysname.tcpsysplexgroupname -
        CLASS(SERVAUTH) ACCESS(READ) ID(userid)
     ```

  4. Use the following command to refresh the SERVAUTH class:

     ```
     SETROPTS RACLIST(SERVAUTH) REFRESH
     ```

  For specific instructions, see the EZARACF sample in SEZAINST.

  Requirements:

  – The AT-TLS policies and SERVAUTH profiles must be coordinated to provide the appropriate level of security. For example:

    - If a connection is presented to the Advisor and the AT-TLS check indicates that this is not a secure connection, then no subsequent SAF check is performed but an IP address ACL check (using the statements in the Load Balancer Advisor configuration files) is performed. If the IP address ACL check is successful, the connection is accepted.
    - If you always want a SAF level check to be performed, then ensure that the AT-TLS policy requires client certificates that are associated with user IDs defined on the local SAF-compliant security product. Also ensure that the SAF profile is defined and the correct user IDs are permitted READ access to the profile. If a secure connection is presented to the Advisor, and the SAF profile exists but the user ID is not authorized to the profile, then the connection will fail and no ACL check will be performed.
    - If the SERVAUTH profile is not defined, then the Load Balancer Advisor performs the IP Address (and port where appropriate) access control list checks before permitting a connection to the Load Balancer Advisor.

  – If the Advisor might run on more than one system, perform set up on all those systems or use a wildcard (*).

  For additional information, see z/OS Security Server RACF Security Administrator's Guide.

Customizing TCP/IP profiles in a CINET environment

The Advisor can use multiple TCP/IP stacks in a CINET environment. The addresses specified in the lb_connection_v4 and lb_connection_v6 statements can belong to different TCP/IP stacks. Moreover, because the socket that listens for Agent connections uses the IPv4 or IPv6 unspecified address, the TCP/IP stack or stacks that incoming Agent connections use depends upon the IP addresses specified in the Agents' agent_id_list statements. To simplify your configuration and to make Advisor outages that are the result of a TCP/IP stack failure or termination more predictable and recoverable, all incoming connections to the Advisor should use a single TCP/IP stack. Therefore, the addresses you specify in the lb_connection_v4 and lb_connection_v6 statements should belong to the same TCP/IP stack, and you should configure all load balancers and Agents to use these same IP addresses when connecting to the Advisor. The addresses you specify should be dynamic VIPAs to enable the movement of the Advisor in case of failure. This implies that these dynamic VIPAs should be defined in the TCP/IP profiles of all the stacks using a VIPARANGE statement. If the Advisor is restarted as a result of failure in a given TCP/IP stack, the dynamic VIPAs are then activated on another TCP/IP stack in that system. If you decide to use the IPv4 or IPv6 unspecified addresses for the lb_connection_v4 and lb_connection_v6 statements, you should use the BIND parameter on the PORT reservation statement to bind these sockets to the dynamic VIPA on the one TCP/IP stack you have decided to use.

Customizing TCP/IP profiles in a subplexing environment

In a subplexing environment, each Advisor and Agent must use a TCP/IP stack that is in its associated subplex. That stack should specify the TCP/IP subplex group ID that corresponds to the TCP/IP part (*tt*) of the sysplex_group_name (EZBT*vvtt*) for which the Advisor or Agent has been configured. The DVIPA for the Advisor must be defined in all the stacks that are associated with the subplex, where a restart of the Advisor can occur.

Step 6: Customize WLM policies for the Advisor and Agents (optional)

It is important that the Advisor and Agents receive an adequate amount of system resources to properly balance workloads. Part of this task involves making the Advisor and Agent run non-swappable. In addition, WLM can control the amount of system resources allocated to the Advisor and Agents.

Guideline: The Advisor and Agents should be assigned to the WLM SYSSTC service class to receive the appropriate dispatching priority. For more information about categorizing work into service classes, see z/OS MVS Planning: Workload Management.

Step 7: Configure the external load balancers

Configure the load balancers with the location (IP address and port) of the Advisor. For maximum availability, this address should be defined as a DVIPA.

In a sysplex subplexing environment, this step requires additional actions. For information about the changes to this step, see “Configuring the external load balancers in a subplexing environment” on page 1170.

If you are using TLS/SSL, the load balancers are client applications in a z/OS Load Balancing Advisor environment. You must configure the load balancers with the same TLS/SSL protocol and cipher suite (if encrypting data) with which the Advisor is configured. Client certificates should be configured on the load balancer, and also defined in the z/OS SAF-compliant security product if that level of authentication is required. See the load balancer documentation for instructions.

Restrictions:

- If the Advisor is using IPv6 for the load balancer connections, or if any Agents are using IPv6 to connect to the Advisor, movement of the Advisor is limited to IPv6–enabled TCP/IP stacks.
- There is a maximum limit of 100 load balancers that can be connected to an Advisor at any given time.

You might be able to customize features of the load balancer's communication with the Advisor. The SASP protocol defines two features that the load balancer implementation might or might not allow to be configured. One determines whether the load balancer polls the Advisor for updated data, or whether

updated data is pushed to the load balancer. The other determines whether only members that have updated data should be sent to the load balancer, or whether all members should be sent to the load balancer regardless of whether their data has changed or not. To determine whether these features can be customized, and how to perform the customization if available, consult your load balancer's documentation. If the load balancer is capable and configured to request that the Advisor push updated information to the load balancer, the Advisor will update the load balancer at least every update interval. If the load balancer is capable and configured to poll the Advisor for updated information, the Advisor will recommend to the load balancer that it poll every update interval. However, the load balancer can choose to disregard this guideline. Consult the load balancer documentation for the expected behavior in these circumstances.

You might want to consider having redundant load balancers configured alike for availability reasons. If so, you need to be aware of the load balancer's unique load balancer identifier (LB UID), sometimes referred to as the UID or UUID, which uniquely identifies a load balancer. Duplicate LB UIDs are not allowed and connection attempts to the Advisor from a load balancer using the same LB UID as an existing connection will force the existing connection to be broken and replaced by the new connection. Redundantly configured load balancers either need to have unique LB UIDs, if you want them to serve as hot standbys that are connected simultaneously along with the load balancer they are backing up, or if they are configured with the same LB UID, they must remain unconnected from the Advisor until the original load balancer fails.

Some load balancers might be capable of using either dispatch or directed mode when forwarding packets to their destinations. External load balancers typically use a cluster IP address to represent the set of applications being load balanced. Client applications use this cluster IP address as the destination IP address for their requests. When a load balancer uses dispatch mode, the destination IP addresses for incoming IP packets is not changed. Instead, the load balancer forwards the packet to a target z/OS system by using the MAC address of a network adapter on that system. The receiving z/OS system inspects the destination IP address of the packet, and if it matches one of the IP addresses in its HOME list, accepts the packet. As a result, with dispatch mode, all target z/OS systems must have the load balancer's cluster IP address defined in their HOME list. However, it is important that these addresses are not advertised externally through dynamic routing protocols. One way to accomplish this is by defining these IP addresses as loopback addresses on z/OS.

With directed mode, the load balancer converts the destination IP address (that is, the cluster IP address) to an IP address owned by the target z/OS system, using technologies such as network address translation (NAT). When IP packets for these connections are sent back to clients, the load balancer converts the source IP address (that is, the target z/OS system's IP address) back to the cluster IP address that the application had used on its request.

While dispatch mode eliminates the need for performing NAT, it does have some special considerations. For example, in Figure 148 on page 1188, both SYSA and SYSB have the same server, FTPD, bound to the same port number using INADDR_ANY. The packet will have the cluster IP address (both SYSA and SYSB are in the same cluster), so the load balancer will use the MAC address to decide to send the packet to SYSA or SYSB, and TCP/IP will then route the packet to the server.

Restrictions: When using dispatch mode, for the load balancer to function correctly, there are the following limitations:

- An OSA can be shared among LPARs only if Virtual MAC (VMAC) addressing is configured for each TCP/IP target stack sharing the OSA.
- All target applications must bind to the IP address specified by INADDR_ANY or in6addr_any, and the cluster IP address must be defined to the stack; however, this must be done so that the address is not advertised (as in a loopback address).

If these restrictions are not met, load balancing will not be optimal because some servers will not get work routed to them.

With directed mode, either the destination IP address (server NAT) is modified in the packet itself, or both the destination and source IP addresses (server NAT and client NAT) are modified in the packet. The packet must return through the same load balancer that will recognize the changes and do the reverse mapping, so a packet can flow from the original destination to the original source.

Configure each load balancer with the members that represent the individual target application instances, or system members that generically represent a system in the sysplex, or both. Members that can share the same type of workload are defined under the same group. For example, TN3270E Telnet servers are defined under one group, and HTTP servers in another. Application members are defined by specifying an IP address, a nonzero port, and a nonzero protocol. System members are defined by specifying an IP address, and specifying the port and protocol to be zero. Members that have only a port of zero or only a protocol of zero (that is, one but not both are zero) are not considered valid members and will not receive any data from the Advisor. The IP addresses of the members must represent valid, reachable addresses within the sysplex that are unique to a specific sysplex system. This excludes such addresses as the loopback addresses, and other non-advertised addresses.

Tip: The Advisor does not check for improperly configured members. After the entire z/OS Load Balancing Advisor system is operational, display all members registered by each load balancer and verify each member you expect to be available is flagged as available. Screen any unavailable members for coding errors in the member, such as incorrect IP addresses, ports, or protocols.

Guideline: For availability reasons, the IP addresses configured for each member should be VIPA addresses (static or dynamic). If the IP address of a physical interface fails and a member specifies that IP address, the Advisor still indicates that the member is available, as alternate routing paths to that member might exist. However, if no alternate routing paths exist, workload requests cannot be delivered to the target system. By using static or dynamic VIPAs in members, the chance of an alternate route being available when a physical interface fails is greatly increased, as long as at least one physical interface is still available.

Restrictions:

- All IP addresses configured in members belonging to the same group must exist within the same sysplex.
- All members belonging to the same group must be of the same type. That is, all members must be application members or all must be system members.
- Certain classes of IP addresses must not be coded for members in the load balancer. This includes the following classes of addresses:
  - Distributed DVIPAs (the address specified on a VIPADISTRIBUTE statement). Defining members with these addresses would combine two load balancing methodologies for the same workload, wasting system resources.
  - Deprecated IPv6 addresses. These are flagged as such in a NETSTAT HOME display. It is probably safest to not code any autoconfigured IPv6 addresses within members.
  - Addresses that are not unique within the sysplex.
  - Addresses that are not reachable from the load balancer, including:
    - Loopback addresses.
    - Unavailable IPv6 addresses. These might be marked as unavailable if duplicate address detection is in progress, has failed, or the interface ID is unknown. These addresses are displayed in a NETSTAT HOME display, including the reason they are marked unavailable.

Configuring the external load balancers in a subplexing environment

You might configure separate load balancers for each subplex, if the subplexes represent connectivity to networks with different security domains. When configuring a load balancer with the IP address of a Load Balancing Advisor, ensure that you have connectivity from the load balancer to the subplex that the Load Balancing Advisor is handling. In addition, groups and target applications that the external load balancer requests information on should belong to the same subplex that the Load Balancing Advisor is handling.

Steps for starting the z/OS Load Balancing Advisor

Follow these steps to start the TCP/IP stacks that the Advisor and Agents are to use, the target applications, the Agents, the Advisor, and the load balancers.

Procedure

Perform the following steps to start the z/OS Load Balancing Advisor:

1. Start the TCP/IP stacks that the Advisor and the Agents will use.
   See "Step 1: Start the TCP/IP stacks that the Advisor and the Agents will use" on page 1171.
2. Start the target applications that will be the targets of load balancing.
   See "Step 2: Start the target applications that will be the targets of load balancing" on page 1171.
3. Start one Agent on each sysplex system that you want to participate in this method of workload balancing.
   See "Step 3: Start one Agent on each sysplex system you want to participate in this method of workload balancing" on page 1171.
4. Start the one instance of the Advisor in the sysplex.
   See "Step 4: Start the one instance of the Advisor in the sysplex" on page 1172.
5. Start the load balancers.
   See "Step 5: Start the load balancers" on page 1172.

Step 1: Start the TCP/IP stacks that the Advisor and the Agents will use

The TCP/IP stacks that the Advisor will use must be started prior to starting the Advisor. An Agent can be started before the TCP/IP stack it uses is started. If the TCP/IP stack that an Agent uses terminates, the Agent remains active and reestablishes communication with the TCP/IP stack after it becomes active again. For CINET considerations regarding Agent recovery after stack failure, see "Starting the TCP/IP stacks in a CINET environment" on page 1171.

Starting the TCP/IP stacks in a CINET environment

In a CINET environment, certain configurations might necessitate manual intervention for recovery if the Agent's TCP/IP stack fails. For more information, see the rule in "Configuring one Agent per z/OS system in the sysplex in a CINET environment" on page 1162.

Step 2: Start the target applications that will be the targets of load balancing

No modifications are necessary to these applications, their configurations, or start procedures, unless the load balancer is using dispatch mode for packet forwarding. For more information on dispatch mode, see "Step 7: Configure the external load balancers" on page 1168.

Step 3: Start one Agent on each sysplex system you want to participate in this method of workload balancing

It does not matter whether the Agents are started before the Advisor, or whether the Advisor is started before the Agents. If the Advisor is started after the Agents are started, the Agent periodically attempts to connect to the Advisor. Only one Agent can be started per z/OS system. Agents must be started from a start procedure as a started program (EXEC PGM=). They cannot be started under BPXBATCH. The IBM-supplied program properties table has entries to make the Agents run non-swappable. You should not override this entry to make the Agents run swappable.

Starting Agents in a subplexing environment

You can start more than one Agent on a z/OS system, if there are TCP/IP stacks for more than one subplex on that system.

Step 4: Start the one instance of the Advisor in the sysplex

As Agents connect to the Advisor, MVS console messages appear on the Advisor's MVS console and on the Agents' MVS consoles. Verify that each Agent you expect to connect to the Advisor has connected. You can also use the NETSTAT,CONN command on the Advisor's TCP/IP stack to see which Agents are currently connected. The Advisor must be started from a start procedure as a started program (EXEC PGM=). It cannot be started under BPXBATCH. The IBM-supplied program properties table has entries to make the Advisor run non-swappable. You should not override this entry to make the Advisor run swappable.

Starting Advisors in a subplexing environment

You can start more than one Advisor in the sysplex, one for each subplex in the sysplex.

Step 5: Start the load balancers

When a load balancer has connected, messages appear on the Advisor's MVS console. You can also use the Advisor's MODIFY *procname*,DISPLAY,LB command to see which load balancers are connected to the Advisor. For details on the Advisor's MODIFY command, see z/OS Communications Server: IP System Administrator's Commands.

Restriction: There is a maximum limit of 100 load balancers that can be connected to an Advisor at any given time.

Verifying that the Advisor system is functioning correctly (optional)

View the MVS console of the Advisor and Agent systems after they are started to verify that the applications started correctly and are still running. If there are any failure messages, see the appropriate message description for the appropriate corrective action. View the syslogd files of the Advisor and Agent systems to see whether any error or warning messages were issued.

Use the following commands to verify that the Advisor system is functioning correctly:

- Verify that the Advisor is started and connected to the expected load balancers by issuing the Advisor's MODIFY *procname*,DISPLAY,LB command. Verify that each load balancer is displayed. Take note of the LB INDEX displayed for each load balancer. This identifier is needed to display details of the load balancer, including its registered groups and members.

  Tip: Individual load balancers are identified in displays by their load balancer index (LB Index), which is generated by the Advisor when the load balancer first connects. To display the details of a particular load balancer, first obtain the LB Index by displaying the list of all load balancers using the Advisor's MODIFY *procname*,DISPLAY,LB command. When you have the LB Index, display the details of a particular load balancer using the appropriate LB Index on the INDEX parameter as follows:

  ```
  MODIFY procname,DISPLAY,LB,INDEX=lb_index
  ```

- Verify that each load balancer configured and registered the appropriate groups and members with the Advisor by issuing the following command:

  ```
  MODIFY procname,DISPLAY,LB,INDEX=assigned_lb_index
  ```

  This display should show all groups and members defined to the load balancer.

- Verify that each Agent has started properly and is communicating with the Advisor. On each Agent, issue the following command:

  ```
  MODIFY procname,DISPLAY,MEMBERS
  ```

  Each member that has an IP address owned by this Agent should appear in the display.

- Verify that the target applications that you want to load balance to are actually available for load balancing. On the Advisor, for each load balancer connected to the Advisor, issue the following command:

  ```
  MODIFY procname,DISPLAY,LB,INDEX=assigned_lb_index
  ```

  Check for the AVAIL flag for each member in the display. The flag is either YES, meaning the member is available for load balancing, or NO, meaning it is not available for load balancing. To be available for load balancing, all of the following conditions must be true:

  – The Agent owning the member's IP address must be active and communicating with the Advisor.
  – The application must be active, if the member represents an application member, and must be on a TCP/IP stack that has not had eventual action message EZD1973E issued by sysplex problem detection and recovery. For more information, see "Problem detection" on page 453.
  – The member must not be quiesced by the Agent operator or the load balancer. The z/OS Agent operator is able to quiesce any member that is owned by that Agent. Also, depending upon load balancer implementation, it might be possible for the load balancer administrator to quiesce individual members.

  If one of the above conditions is false, correct the situation and repeat the display command until you are satisfied that all members that you intend to have available for load balancing are displayed as being available.
- Verify that the Advisor system is functioning correctly when using AT-TLS:
  – Use the **pasearch** command from the z/OS UNIX shell to query information from the Policy Agent. For example, **pasearch -t -r** displays active AT-TLS rule details. For more information about displaying policy based networking information, see z/OS Communications Server: IP System Administrator's Commands.
  – Use the Netstat TTLS/-x command to display z/OS Load Balancing Advisor, Agent, and ADNR AT-TLS policies. For more information about the Netstat TTLS/-x report, see z/OS Communications Server: IP System Administrator's Commands.

Operating the z/OS Load Balancing Advisor

When the Advisor and Agent are operational, you can monitor and customize processing:

- Change the logging level of the Advisor and Agents to suit your needs (optional)
- Interpret Agent and Advisor display information
- Stop distributing new workload requests (QUIESCE) to particular members or resume distribution (ENABLE) to those members (optional)

Changing the logging level of the Advisor and Agents

Optionally, you can change the logging level of the Advisor and Agents to suit your needs. The amount of information that is logged by the Advisor and Agents can be modified dynamically using the following command:

```
MODIFY procname,DEBUG,LEVEL=debug_level
```

However, modifying the logging level is not a step that should be taken lightly. The IBM default of 7 should not be changed unless instructed to do so by an IBM Service representative. For things to consider before modifying this value, see "Customizing optional statements" on page 1158.

Interpreting Agent and Advisor display information

Successfully interpreting the various flags and indicators that appear in Advisor and Agent displays can help you identify configuration problems, or might help explain why workloads are not being distributed as expected. For information on each field in the Advisor and Agent displays, see z/OS Communications

Server: IP System Administrator's Commands. Some portions of the displays are described in more depth in Table 61 on page 1174, Table 62 on page 1177, and the subtopics that follow.

Table 61. Summary of selected Advisor display output fields and flags

Flag or field name	Location	Description
ABNORM	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	The rate of abnormal transaction completions per 1000 total transaction completions for this application. This value is optionally provided to WLM by the application. WLM provides this value along with the server-specific recommendation on the system where the member is. For 1000 total transactions completed, it displays the number of those transactions that could not successfully complete. If the value is nonzero, WLM uses it to reduce the server-specific recommendation (WLM weight).
AVAIL	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	Indicates whether the member is available for workload balancing requests.
BASEWLM	GROUP FLAGS field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	WLM system weight recommendations were configured or specified by default for the members of this group, and are being used as a component of the net weight.
BASEWLM*	GROUP FLAGS field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=nn command output	Server-specific WLM recommendations were configured or specified by default for the members of this group. However, WLM system weight recommendations are actually being used as a component of the net weight.
CP	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	When shown on the next line after ProcType, indicates the proportion that is applied against the CP weight. When shown on the line starting with RAW, indicates the raw CP weight that was returned by WLM. When shown on the line starting with Proportional, indicates the proportionally adjusted raw CP weight that was used to determine the BASEWLM or SERVERWLM composite weight.
CS WEIGHT	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	Communications Server weight. A value calculated by the Agent that owns the member, representing the health of the application with respect to its ability to process the work that has recently been received.

Table 61. Summary of selected Advisor display output fields and flags (continued)

Flag or field name	Location	Description
HEALTH	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	The general health of the application. This value is optionally provided to WLM by the application. WLM provides this value along with the server-specific recommendation on the system where the member is. If the value is less than 100, WLM uses it to reduce the server-specific recommendation (WLM weight).
LB INDEX	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=*nn* commands, an output field of the load balancer area	A unique identifier assigned to a load balancer for the purpose of referencing the load balancer in subsequent operator commands
LBQ	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	The load balancer has quiesced the member.
NET WEIGHT	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	The weight the Advisor passes to the load balancer, representing the desirability of the member to receive additional workload requests relative to the other members of the same group. The net weight is calculated using the WLM weight and Communications Server weight.
NOCHANGE	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output, the FLAGS field of the load balancer area	The load balancer has requested that it be sent only information about members that have changed their status or weights since the last time the load balancer received information on its members.
NODATA	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	A transient flag indicating that the responsible Agent has not had enough time to calculate a Communications Server weight.
NOTARGETAPP	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	The target application of this application member is not active.
NOTARGETIP	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	This is an unusable system member.
NOTARGETSYS	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	The Advisor is not aware of the system that owns the IP address of the member.
OPQ	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	The z/OS operator has quiesced the member.

Table 61. Summary of selected Advisor display output fields and flags (continued)		
Flag or field name	**Location**	**Description**
ProcType	A field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	Indicates that CP, zAAP, and zIIP proportions were configured. These proportions are applied against the processor weights to determine the composite BASEWLM weight.
PUSH	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output, the FLAGS field of the load balancer area	The load balancer has requested to receive information from the Advisor on a scheduled basis, rather than having to poll the Advisor for the information.
SERVERWLM	GROUP FLAGS field of the group area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	Server-specific WLM recommendations are being used for the members of this group as a component of the net weight.
TRUST	For the MODIFY *procname*,DISPLAY,LB and MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output, the FLAGS field of the load balancer area	The load balancer will allow other system components besides itself to register members with the load balancer. The z/OS Load Balancing Advisor does not currently exploit this feature.
WLM WEIGHT	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	WorkLoad Manager weight. The value received from WLM on the system where the member is, representing the displaceable capacity of that system relative to other systems in the sysplex (system weight), or the value received from WLM representing how well the server is performing relative to its WLM policies (server-specific weight). This weight is a composite weight determined from the displayed CP, zAAP, and zIIP weights.
zAAP	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	When shown on the next line after ProcType, indicates the proportion that is applied against the zAAP weight. When shown on the line starting with RAW, indicates the raw zAAP weight that was returned by WLM. When shown on the line starting with Proportional, indicates the proportionally adjusted raw zAAP weight that was used to determine the BASEWLM or SERVERWLM composite weight.

Table 61. Summary of selected Advisor display output fields and flags (continued)		
Flag or field name	**Location**	**Description**
zIIP	A field of the member area of the MODIFY *procname*,DISPLAY,LB,INDEX=*nn* command output	When shown on the next line after ProcType, indicates the proportion that is applied against the zIIP weight. When shown on the line starting with RAW, indicates the raw zIIP weight that was returned by WLM. When shown on the line starting with Proportional, indicates the proportionally adjusted raw zIIP weight that was used to determine the BASEWLM or SERVERWLM composite weight.

Table 62. Summary of selected Agent display output flags		
Flag or field name	**Location**	**Description**
ANY	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,MEMBERS command output	The target application is bound to the unspecified address, 0.0.0.0 for IPv4 or :: for IPv6.
V6	FLAGS field of the member area of the MODIFY *procname*,DISPLAY,MEMBERS command output	The target application was bound using the IPV6_V6ONLY socket option.

MODIFY *procname*,DISPLAY,LB

This command displays all load balancers that are currently connected to the Advisor. In Figure 144 on page 1177, the line numbers appearing in the left margin are for reference purposes only, and are used in the subtopics following the figure.

```
1  MODIFY LBADV,DISPLAY,LB
2  EZD1242I LOAD BALANCER SUMMARY
3  LB INDEX      : 00        UUID      : 637FFF175C
4   IPADDR..PORT : 10.42.154.105..50005
5   HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
6  LB INDEX      : 01        UUID      : 207FFF175C
7   IPADDR..PORT : 10.42.105.60..50006
8   HEALTH       : 7F        FLAGS     : PUSH TRUST
9  2 OF 2 RECORDS DISPLAYED
```

Figure 144. Sample output for the MODIFY procname,DISPLAY,LB command

LB INDEX

The LB index is generated by the Advisor to uniquely identify a load balancer, and is used in other MODIFY commands to display details of a particular load balancer. The LB index that is assigned to the next load balancer that connects can be difficult to predict at times, in case your automation attempts to predict them. Each time a load balancer connects, it is assigned a different LB index. At first, the numbers are assigned in order from 0 to 99, and then the numbers are reused. The next number assigned after 99 is based on a least-recently-used algorithm. Thus, of the load balancer connections that used the numbers 0-99, the index of the load balancer connection that ended first would be the next number assigned as the next LB index. This is done to prevent the confusion that could result if new load balancer connections obtained the lowest available index. If that were the case, it might be difficult to ascertain whether different load balancer displays referred to the same or different load balancer connection instances. If the Advisor is brought down and back up, the indexes start from zero again. Lines 3 and 6 in Figure 144 on page 1177 show two LB indexes. For more information about the LB INDEX field, see z/OS Communications Server: IP System Administrator's Commands.

NOCHANGE, PUSH, TRUST

All of these flags are set only by the load balancer. Whether they appear or not depends upon whether the particular load balancer implementation supports them, and if they are configured in the load balancer. The z/OS administrator has no control over the settings of these flags.

The NOCHANGE flag can affect the amount of data transferred between the load balancer and the Advisor. If this flag is set, only data that has changed since the last time data was sent to the load balancer is included. Consult the load balancer documentation to determine whether this setting is supported.

The PUSH flag can have an effect on how soon the load balancer is informed of certain events. If the PUSH flag is not on, the load balancer must poll the Advisor periodically for updates. If this flag is on, certain events can be communicated to the load balancer earlier than would be possible if polling were in effect. Those events include quicker notification of a target application being taken out of service, and quicker notification of when the member's IP address has been moved to another system in the sysplex (VIPA takeover) or removed entirely. When the PUSH flag is on, the Advisor also sends the load balancer updated information about weights and status at least every update interval, if new information is available. Consult the load balancer documentation to determine whether this setting is supported.

The TRUST flag indicates that the load balancer allows other system components besides itself to register members with the load balancer. The z/OS Load Balancing Advisor does not currently exploit this feature.

Line 5 of Figure 144 on page 1177 shows all of these flags. For more information on the NOCHANGE, PUSH, and TRUST flags, see z/OS Communications Server: IP System Administrator's Commands.

MODIFY *procname*,DISPLAY,LB,INDEX=*lbindex*

This command displays details about a particular load balancer, including its registered groups and members. In Figure 145 on page 1179, the line numbers appearing in the left margin are for reference purposes only, and are used in the subtopics following the figure.

```
1    MODIFY LBADV,DISP,LB,INDEX=0
2    EZD1243I LOAD BALANCER DETAILS
3    LB INDEX      : 00        UUID      : 637FFF175C
4   IPADDR..PORT : 10.42.154.105..50005
5   HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
6   GROUP NAME   : SYSTEMFARM
7    GROUP FLAGS : BASEWLM
8    IPADDR..PORT: 10.42.105.154..0
9     SYSTEM NAME: MVS209    PROTOCOL  : 000  AVAIL     : YES
10    WLM WEIGHT : 00040     CS WEIGHT : 100  NET WEIGHT: 00001
10a     RAW           CP: 40  zAAP: 60  zIIP: 00
10b     Proportional  CP: 40  zAAP: 00  zIIP: 00
11    FLAGS      :
12   IPADDR..PORT: 10.42.105.60..0
13    SYSTEM NAME: VIC007    PROTOCOL  : 000  AVAIL     : YES
14    WLM WEIGHT : 00050     CS WEIGHT : 100  NET WEIGHT: 00002
14a     RAW           CP: 50  zAAP: 00  zIIP: 00
14b     Proportional  CP: 00  zAAP: 00  zIIP: 00
15    FLAGS      :
16   IPADDR..PORT: 10.42.105.22..0
17    SYSTEM NAME: N/A       PROTOCOL  : 000  AVAIL     : NO
18    WLM WEIGHT : 00000     CS WEIGHT : 000  NET WEIGHT: 00000
18a     RAW           CP: 00  zAAP: 00  zIIP: 00
18b     Proportional  CP: 00  zAAP: 00  zIIP: 00
19    FLAGS      : NOTARGETSYS
20   IPADDR..PORT: 10:1::4:5..0
21    SYSTEM NAME: MVS209    PROTOCOL  : 000  AVAIL     : NO
22    WLM WEIGHT : 00040     CS WEIGHT : 000  NET WEIGHT: 00000
22a     RAW           CP: 40  zAAP: 60  zIIP: 00
22b     Proportional  CP: 40  zAAP: 00  zIIP: 00
23    FLAGS      : NOTARGETIP
24  GROUP NAME   : UDP_SERVER_FARM
25   GROUP FLAGS : SERVERWLM
26   IPADDR..PORT: 10.42.105.154..7777
27    SYSTEM NAME: MVS209    PROTOCOL  : UDP  AVAIL     : YES
28    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
28a     RAW           CP: 20  zAAP: 22  zIIP: 00
28b     Proportional  CP: 10  zAAP: 11  zIIP: 00
28c     ABNORM     : 00200     HEALTH    : 100
29    FLAGS      :
30   IPADDR..PORT: 2001:DB8::10:5:6:2..7777
31    SYSTEM NAME: MVS209    PROTOCOL  : UDP  AVAIL     : YES
32    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
32a     RAW           CP: 25  zAAP: 18  zIIP: 00
32b     Proportional  CP: 10  zAAP: 11  zIIP: 00
33    FLAGS      :
34   IPADDR..PORT: 10.42.105.60..7777
35    SYSTEM NAME: VIC007    PROTOCOL  : UDP  AVAIL     : YES
36    WLM WEIGHT : 00045     CS WEIGHT : 100  NET WEIGHT: 00002
36a     RAW           CP: 50  zAAP: 18  zIIP: 00
36b     Proportional  CP: 30  zAAP: 15  zIIP: 00
```

Figure 145. Sample output for the MODIFY procname,DISPLAY,LB,INDEX=lbindex command, part 1 of 2

```
37    FLAGS      :
38  GROUP NAME   : DNS_GROUP
39   GROUP FLAGS : BASEWLM*
40   IPADDR..PORT: 10.42.103.75..53
41    SYSTEM NAME: MVSW      PROTOCOL  : TCP  AVAIL     : NO
42    WLM WEIGHT : 00064     CS WEIGHT : 100  NET WEIGHT: 00000
42a     RAW           CP: 64  zAAP: 00  zIIP: 00
42b     Proportional  CP: 64  zAAP: 00  zIIP: 00
43    FLAGS      : LBQ OPQ
44   IPADDR..PORT: 10.42.105.60..53
45    SYSTEM NAME: VIC007    PROTOCOL  : TCP  AVAIL     : NO
46    WLM WEIGHT : 00050     CS WEIGHT : 000  NET WEIGHT: 00000
46a     RAW           CP: 50  zAAP: 00  zIIP: 00
46b     Proportional  CP: 50  zAAP: 00  zIIP: 00
47    FLAGS      : NOTARGETAPP
48   IPADDR..PORT: 10.42.105.154..53
49    SYSTEM NAME: MVS209    PROTOCOL  : TCP  AVAIL     : YES
50    WLM WEIGHT : 00040     CS WEIGHT : 100  NET WEIGHT: 00021
50a     RAW           CP: 40  zAAP: 00  zIIP: 00
50b     Proportional  CP: 40  zAAP: 00  zIIP: 00
51    FLAGS      : NODATA
52  GROUP NAME   : CICS_SERVER_FARM
53   GROUP FLAGS : BASEWLM
54   ProcType    :
54a    CP : 060  zAAP: 040  zIIP: 000
55   IPADDR..PORT: 10.42.154.105..8888
56    SYSTEM NAME: MVS209    PROTOCOL  : TCP  AVAIL     : YES
57    WLM WEIGHT : 00048     CS WEIGHT : 100  NET WEIGHT: 00001
57a     RAW           CP: 40  zAAP: 60  zIIP: 00
57b     Proportional  CP: 24  zAAP: 24  zIIP: 00
58    FLAGS      :
59   IPADDR..PORT: 10.42.105.60..8888
60    SYSTEM NAME: VIC007    PROTOCOL  : TCP  AVAIL     : YES
61    WLM WEIGHT : 00054     CS WEIGHT : 100  NET WEIGHT: 00001
61a     RAW           CP: 50  zAAP: 60  zIIP: 00
61b     Proportional  CP: 30  zAAP: 24  zIIP: 00
62    FLAGS      :
63   IPADDR..PORT: 10.42.105.22..8888
64    SYSTEM NAME: N/A       PROTOCOL  : TCP  AVAIL     : NO
65    WLM WEIGHT : 00000     CS WEIGHT : 000  NET WEIGHT: 00000
65a     RAW           CP: 00  zAAP: 00  zIIP: 00
65b     Proportional  CP: 00  zAAP: 00  zIIP: 00
66    FLAGS      : NOTARGETSYS
67   IPADDR..PORT: 10:1::4:5..8888
68    SYSTEM NAME: MVS209    PROTOCOL  : TCP  AVAIL     : NO
69    WLM WEIGHT : 00048     CS WEIGHT : 000  NET WEIGHT: 00001
69a     RAW           CP: 40  zAAP: 60  zIIP: 00
69b     Proportional  CP: 24  zAAP: 24  zIIP: 00
70    FLAGS      : NOTARGETIP
71 14 OF 14 RECORDS DISPLAYED
```

Figure 146. Sample output for the MODIFY procname,DISPLAY,LB,INDEX=lbindex command, part 2 of 2

Group flags - BASEWLM, BASEWLM, and SERVERWLM*

The BASEWLM flag indicates that WLM system weight recommendations were configured or specified by default for this group, and are being used to calculate the net weight. The BASEWLM* flag indicates that server-specific WLM recommendations were configured for this group, but WLM system weight recommendations are being used instead. This occurs when at least one of the Agents owning members within the group does not support server-specific WLM recommendations. The SERVERWLM flag indicates that server-specific WLM recommendations are being used to calculate the net weight for each member in the group. Line 7 in Figure 145 on page 1179 shows the BASEWLM flag, line 39 shows the BASEWLM* flag, and line 25 shows the SERVERWLM flag. For more information on the BASEWLM, BASEWLM*, and SERVERWLM flags, see z/OS Communications Server: IP System Administrator's Commands.

Member flags - LBQ and OPQ

The LBQ flag indicates that the load balancer has quiesced the member. For details on what this entails, see “Stopping or resuming workload distribution to particular members (QUIESCE and ENABLE)” on page 1186. Do not confuse this with the OPQ flag, which indicates that the z/OS operator has quiesced the member at the z/OS Agent. In both cases, the member is ineligible for future workloads through the external load balancer. Line 43 in Figure 145 on page 1179 shows the LBQ flag and the OPQ flag. For

more information on the LBQ and OPQ flags, see z/OS Communications Server: IP System Administrator's Commands.

Member flags - NOTARGETSYS, NOTARGETIP, and NOTARGETAPP

These flags indicate that the Advisor will advise the load balancer that the member should not currently receive new workload requests because a resource is unavailable. If the IP address in the member is not present on any TCP/IP stack within the sysplex or subplex, NOTARGETSYS is displayed for that member. This flag can also appear if the Agent owning the IP address in the member has lost contact with the Advisor or has yet to be started. There might be rare instances where the load balancer might decide to go ahead and route workload requests to members that have the NOTARGETSYS flag displayed, if it has no better candidates within the group to route workload requests to. If the member represents an application member and the application is not active, NOTARGETAPP is displayed for that member. If the member is a system member and the IP address is unusable, NOTARGETIP is displayed. This includes distributed VIPAs (DVIPAs), deprecated IPv6 addresses, and unavailable IPv6 addresses. If you ever see the NOTARGETIP flag, you should update the IP address in the member at the load balancer. If application members are coded with any of these addresses, you will always see NOTARGETAPP displayed for these members. Line 19 in Figure 145 on page 1179 shows the NOTARGETSYS flag, line 23 shows the NOTARGETIP flag, and line 47 shows the NOTARGETAPP flag. For more information on the NOTARGETSYS, NOTARGETIP, and NOTARGETAPP flags, see z/OS Communications Server: IP System Administrator's Commands.

Member flag - NODATA

This flag is transient and indicates that not enough time has elapsed for the reporting Agent to calculate a Communications Server weight for the member. Therefore, only the WLM weight is used to calculate the net weight, until such time that the Agent can report a Communications Server weight. Until that time, the CS WEIGHT is displayed as 100. When a Communications Server weight has been calculated and transmitted to the Advisor, the NODATA flag is turned off. This flag appears when new members are registered by the load balancer, when the target application that the member represents first becomes active, or shortly after a target application has been moved within the sysplex or subplex. Line 51 in Figure 145 on page 1179 shows the NODATA flag. For more information on the NODATA flag, see z/OS Communications Server: IP System Administrator's Commands.

Member field - AVAIL

This field is always displayed for each registered member, and has a value of YES or NO. YES indicates that the member is available for workload request distribution. NO indicates that the member is not available for workload request distribution. For the member to be available for workload request distribution, the target application must be active (if the member represents an application member) on a TCP/IP stack that has not had eventual action message EZD1973E issued by sysplex problem detection and recovery, an Agent must be active on the target system and connected to the Advisor, and the member must not be quiesced by the z/OS operator or by the load balancer. Line 9 in Figure 145 on page 1179 shows an example of the AVAIL field set to YES, and line 17 shows an example of the AVAIL field set to NO. For more information on the AVAIL field, see z/OS Communications Server: IP System Administrator's Commands.

Member field - NET WEIGHT

This field is always present for each registered member, and is the only weight that the load balancer actually receives. It is calculated by applying the Communications Server weight as a percentage of the WLM weight, and then the weight is normalized within its group. Normalization involves reducing the weight values while largely preserving the ratios between the weights. Normalization is performed within a group only if there is more than one available member within the group. Net weights can be in the range 0 – 64. A higher net weight means that member is capable of receiving more work than a member within the same group that has a lower weight. There are certain situations where the net weight can be 0 when neither the WLM weight or the Communications Server weight is 0. This can happen if the member has been quiesced by the z/OS operator or the load balancer. Conversely, there is one case where the WLM weight or the Communications Server weight can be zero, but the net weight is nonzero. This can happen if the net weight of every member in the group calculates to zero, and at least one member of the group is available. In this case, the net weights of all of the available members in the group are changed to 1 to force round-robin distribution among the members in the group, rather than to stop sending new workloads to the group entirely. Line 10 in Figure 145 on page 1179 shows an example of the NET WEIGHT field set to a nonzero value, while line 18 shows an example of the field set to zero. For more information about the NET WEIGHT field, see z/OS Communications Server: IP System Administrator's Commands.

Member field - WLM WEIGHT

This field is always present for each registered member, and represents the desirability of the system owning the member, relative to the other systems in the sysplex or subplex (system weight), or a measure of how well the individual application is meeting its WLM policies (server-specific weight). Like the Communications Server weight and net weight, a higher value means it is more desirable. WLM weights can be in the range 0 – 64. The WLM weight is used as a key component of the net weight. The WLM weight is the composite weight, and is the sum of the modified CP, zAAP, and zIIP weights displayed on the next line. Line 10 in Figure 145 on page 1179 is one of many lines that contain the WLM WEIGHT field, and Lines 10a and 10b are some of the many lines that contain the modified CP, zAAP, and zIIP weights described in Table 63 on page 1183.

Table 63. WLM WEIGHT - CP, zAAP, and zIIP fields

Processor	DISTMETHOD BASEWLM	DISTMETHOD SERVERWLM
CP	The first value is the WLM system general CP weight recommendation. It is based on the amount of displaceable general CPU capacity on this system, as compared to the other target systems. The second value is the first value modified by the expected general CP usage proportion configured on the PORTLIST and WLM statement for this application.	The first value is the WLM server-specific general CP recommendation. This is the amount of displaceable general CPU capacity, based on the application workload's importance (as defined by the WLM policy) as compared to the other target systems. The second value is the first value modified by the proportion of general CP capacity that is currently being consumed by the application's workload, as compared to the other processors (zAAP and zIIP).
zAAP	The first value is the WLM system zAAP weight recommendation. It is based on the amount of displaceable zAAP capacity on this system, as compared to the other target systems. The second value is the first value modified by the expected zAAP usage proportion configured on the PORTLIST and WLM statement for this application.	The first value is the WLM server-specific zAAP recommendation. This is the amount of displaceable zAAP capacity, based on the importance (as defined by the WLM policy) of the application's workload, as compared to the other target systems. The second value is the first value modified by the proportion of zAAP capacity that is currently being consumed by the application's workload, as compared to the other processors (general CPU and zIIP).
zIIP	The first value is the WLM system zIIP weight recommendation. It is based on the amount of displaceable zIIP capacity on this system, as compared to the other target systems. The second value is the first value modified by the expected zIIP usage proportion configured on the PORTLIST and WLM statement for this application.	The first value is the WLM server-specific zIIP recommendation. This is the amount of displaceable zIIP capacity, based on the importance (as defined by the WLM policy) of the application's workload, as compared to the other target systems. The second value is the first value modified by the proportion of zIIP capacity that is currently being consumed by the application's workload, as compared to the other processors (general CPU and zAAP).

Restriction: This information is available to be displayed only if no systems in the sysplex are prior to z/OS V1R9. If any systems in the sysplex are not at this release level, only CP weights are considered when determining a composite weight recommendation.

For more information about the WLM WEIGHT field, see z/OS Communications Server: IP System Administrator's Commands.

Member field - CS WEIGHT

This CS WEIGHT field is always present for each registered member, and represents the health of the server with respect to its ability to satisfy recent requests. As with the WLM weight and the net weight, the higher the value the better the health. The Communications Server weight can range from 0 to 100, and is used as a component of the net weight. Line 10 in Figure 145 on page 1179 is one of many lines that contains the CS WEIGHT field. For more information on the CS WEIGHT field, see z/OS Communications Server: IP System Administrator's Commands.

Member field - ABNORM

This field is displayed if the GROUP FLAGS field indicates that server-specific (SERVERWLM) WLM recommendations are being used. The ABNORM value is a nonzero value if the server application is experiencing conditions in which transactions are completing abnormally, and represents a rate of abnormal transaction completions per 1000 total transaction completions. It is applicable only for target applications that act as Subsystem Work Managers, reporting transaction status using Workload Management Services, such as IWMRPT. For example, the value of 200 in this example (see Line 28c in Figure 145 on page 1179) indicates that 20% of all transactions processed by the server application are completing abnormally. Under normal conditions or if the server is not providing this information to WLM, this value is 0.

A nonzero value indicates that the server application has reported abnormal transaction completions to WLM and that WLM has reduced the server-specific recommendation for this server instance. The higher the value of this field, the greater the reduction in the recommendation provided by WLM. For more information regarding the conditions leading to abnormal transaction completions for a given server application, see the documentation provided by the server application.

Restriction: Although WLM uses the abnormal transaction completion rate provided by the application to reduce the server-specific recommendation, this information is available for display on an Advisor only if the Load Balancing Agents and the Advisor are running on a z/OS V1R8 system. A z/OS V1R7 Load Balancing Agent does not provide this information to the Load Balancing Advisor. In this situation, a z/OS V1R8 Advisor shows a normal abnormal transaction completion rate of 0, even if WLM is reducing the server-specific recommendation because of a nonzero abnormal transaction completion rate reported from the application.

Member field - HEALTH

This field is displayed if the GROUP FLAGS field indicates that server-specific (SERVERWLM) WLM recommendations are being used. This health indicator is available only for applications that provide this information to WLM using the IWM4HLTH or IWMSRSRG services, and it indicates the general health for an application or subsystem. Under normal circumstances, or if the server is not providing this information to WLM, the value of this field is 100, meaning the server is 100% healthy.

Any value less than 100 indicates that the server is experiencing problem conditions that are preventing it from processing new work requests successfully, which causes WLM to reduce the server-specific recommendation for this server instance. The lower the value of this field, the greater the reduction in the recommendation provided by WLM.

Restriction: Although WLM uses the health indicator provided by the application to reduce the server-specific recommendation, this information is available for display on an Advisor only if the Load Balancing Agents and the Advisor are running on a z/OS V1R8 system. A z/OS V1R7 Load Balancing Agent does not provide this information to the Load Balancing Advisor. In this situation, a z/OS V1R8 Advisor shows a normal health indicator of 100, even if WLM is reducing the server-specific recommendation because of an abnormal health indication from the application.

Member field - ProcType

Indicates that CP, zAAP, and zIIP proportions were configured. These proportions are applied against the processor weights to determine the composite BASEWLM weight. For a description of the CP, zAAP, and zIIP fields for ProcType, see Table 61 on page 1174. Lines 54 and 54a in Figure 145 on page 1179 contain these fields.

MODIFY *procname*,DISPLAY,MEMBERS,DETAIL

This command displays details about members that are owned by the Agent. In Figure 147 on page 1185, the line numbers appearing in the left margin are for reference purposes only, and are used in the subtopics following the figure. For information about every field displayed by the Agent MODIFY command, see z/OS Communications Server: IP System Administrator's Commands.

```
1  MODIFY LBAGENT,DISPLAY,MEMBER,DETAILS
2  EZD1245I MEMBER DETAILS
3  LB INDEX      : 00        UUID      : 637FFF175C
4   GROUP NAME   : SYSTEMFARM
5    GROUP FLAGS : BASEWLM
6    IPADDR..PORT: 10.42.105.154..0
7     TCPNAME    : TCPCS     MATCHES   : 001  PROTOCOL  : 000
8     FLAGS      :
9     JOBNAME    : N/A       ASID      : N/A  RESOURCE  : N/A
10    IPADDR..PORT: 10:1::4:5..0
11    TCPNAME    : TCPCS5    MATCHES   : 000  PROTOCOL  : 000
12    FLAGS      :
13    JOBNAME    : N/A       ASID      : N/A  RESOURCE  : N/A
14  GROUP NAME   : UDP_SERVER_FARM
15   GROUP FLAGS : SERVERWLM
16   IPADDR..PORT: 10.42.105.154..7777
17    TCPNAME    : TCPCS     MATCHES   : 001  PROTOCOL  : UDP
18    FLAGS      : ANY
19    JOBNAME    : TESTD1    ASID      : 0035 RESOURCE  : 000000A3
20   IPADDR..PORT: 2001:DB8::10:5:6:2..7777
21    TCPNAME    : TCPCS2    MATCHES   : 001  PROTOCOL  : UDP
22    FLAGS      : ANY V6
23    JOBNAME    : TESTD2    ASID      : 002A RESOURCE  : 00000031
24 4 OF 4 RECORDS DISPLAYED
```

Figure 147. Sample output for the MODIFY procname,DISPLAY,MEMBERS,DETAIL command

Member flag - ANY

The ANY flag means that the application represented by the port in the member is bound to INADDR_ANY or the IPv6 unspecified address (in6addr_any). This means that in an INET (one TCP/IP stack) environment, any externally available IP address owned by that TCP/IP stack might be used to reach the target application, not just the IP address coded in this member. Therefore, there is the potential that multiple members might exist in the load balancer or other load balancers that actually represent the same application, if members were coded with the same port, protocol, and an IP address owned by the same TCP/IP stack. You need to be aware of this if you want to issue operator commands to quiesce that application. If this were the case, quiescing the application at the port level, but specifying the individual IP address of the member, might not quiesce all new workload requests to that application. Quiescing the application at the port level without specifying an IP address would be required to accomplish that task. If the application is running in a CINET (multiple TCP/IP stack) environment, any externally available IP address on the z/OS system can be used to reach the target application, unless the application has stack affinity. If the application has stack affinity, the Advisor indicates the member is available only if the IP address coded in the member belongs to the TCP/IP stack that the application has affinity to. Line 18 in Figure 147 on page 1185 shows an example of this flag. For more information on the ANY flag, see z/OS Communications Server: IP System Administrator's Commands.

Member flag - V6

The V6 flag indicates that the application that the member refers to has set the IPV6_V6ONLY socket option. This socket option disallows connections to the server application using an IPv4 address as the destination IP address when the server application has bound to the IPv6 unspecified address (in6addr_any). If a member is coded with an IPv4 address and intends to represent an application that

has the IPV6_V6ONLY socket option set, the member will not be available for workload balancing and the V6 flag is displayed for this member. Conversely, for any member that represents this target application and is coded with an IPv6 address that can be used to reach the target application, the member will be available for workload balancing and the V6 flag is displayed. Line 22 in Figure 147 on page 1185 shows an example of this flag. For more information on the V6 flag, see z/OS Communications Server: IP System Administrator's Commands.

Stopping or resuming workload distribution to particular members (QUIESCE and ENABLE)

At least one and possibly two methods exist to stop sending new workload requests to particular members, referred to as quiescing. Quiescing does not disrupt existing connections with the target applications, but does prevent new workload requests from being distributed to those members from load balancers. Requests sent directly to applications that do not go through the load balancer are unaffected by quiescing. Quiescing certain members can be useful for a planned outage of a particular system in the sysplex, a particular TCP/IP stack, a particular application, or a homogeneous group of applications, such as all HTTP servers. The first method of quiescing is done using the MODIFY *procname*,QUIESCE operator command available at each Agent. The second potential method is through the load balancer administrator, if the load balancer implementation supports this function. Only the MODIFY *procname*,QUIESCE command is described in detail in this topic.

The MODIFY *procname*,QUIESCE command is available only on the Agents, because they own the IP addresses that belong to TCP/IP stacks on that z/OS system. Therefore, the scope of the quiesce operation cannot affect members that are not owned by the Agent that is issuing the command.

There are three major scopes or target options for the MODIFY *procname*,QUIESCE command:

- SYSTEM

  Every member owned on that z/OS system can be quiesced
- TCPNAME=*tcpname*

  Every member on one of the Agent's TCP/IP stacks can be quiesced
- PORT=*portnum*

  All members using a particular port can be quiesced

The last target option, quiescing by port, enables you to refine the quiesce to an individual member instead of quiescing all members using the port. For more information on how to specify individual members using the Agent's MODIFY *procname*,QUIESCE command, see z/OS Communications Server: IP System Administrator's Commands.

When ready for new workloads, use the MODIFY *procname*,ENABLE command to make the quiesced members available again. Like the quiesce command, this command is only an Agent command, and also has the same target options as the quiesce command.

The quiesce and enable commands are hierarchical. The system level (SYSTEM target option) is at the top of the hierarchy, the stack level (TCPNAME=*tcpname* target option) is the next highest, and the member level (PORT=*portnum* target option) is the lowest.

Rules:

- In the quiesce and enable hierarchy, a member quiesced at one level of the hierarchy cannot be enabled at a lower or higher level of the hierarchy. For example, if the MODIFY *procname*,QUIESCE,TCPNAME=*tcpname* command has been issued, a MODIFY *procname*,ENABLE,SYSTEM command or a MODIFY *procname*,ENABLE,PORT=*portnum* command will not be accepted.
- An enable command must be issued at the same level of the hierarchy as the last quiesce command that affected the member.
- When a member has been quiesced at one level of the hierarchy, it can be quiesced at a higher level of the hierarchy. This promotes the quiesce level of the member from the lower hierarchy level to the

higher hierarchy level, and any history of it being quiesced at the lower level of the hierarchy is erased. When subsequently issuing an enable command in these circumstances, the enable command must be issued at the higher, promoted level of the hierarchy. For example, if member A was quiesced at the port level, a subsequent command to quiesce all members of the TCP/IP stack would be accepted. Furthermore, only a subsequent enable command at the TCP/IP stack level would be accepted to re-enable the member.

Table 64 on page 1187 shows which quiesce and enable commands are valid for a member after a previous quiesce command has affected that same member. A dot at the intersection of the prior command column and the current command row indicates that the current command would be valid for that member. Absence of a dot indicates that the current command would not be valid after the prior command had affected that member.

Table 64. Allowed quiesce and enable command sequences for members

		Prior command		
		QUIESCE,SYSTEM	QUIESCE,TCPNAME=	QUIESCE,PORT=
Current command	QUIESCE,SYSTEM		●	●
	QUIESCE,TCPNAME=			●
	QUIESCE,PORT=			
	ENABLE,SYSTEM	●		
	ENABLE,TCPNAME=		●	
	ENABLE,PORT=			●

Rules:

- A quiesce command is rejected if any member it applies to has already been quiesced by a command at a higher level of the hierarchy. For example, if you are running in a CINET configuration and you quiesce all members under stack A, which includes a member on stack A that used port 80, you cannot subsequently issue a quiesce command at the port level hoping to quiesce members using port 80 on stacks B and C, because the member on stack A that used that port was already quiesced at the stack level. Because the command would fail for one member, the entire command fails for all members.
- Quiesce commands at the system and stack level apply to currently registered members and also members that are registered in the future, provided that the IP addresses of those future members are owned by the Agent that issued the command. For quiesce commands issued at the stack level, the specified stack must exist at the time the command is issued.
- Quiesce commands at the port level can apply to members registered in the future, if a member currently exists at that port number. For example, if a member is registered by one load balancer at port 80 and the Agent operator quiesces all members at port 80, and then another load balancer registers that same member (same IP address, port, and protocol), the newly registered member would inherit the quiesce performed at the port level.

When a member represents a shareport group (that is, multiple server application instances sharing the same TCP port on the same TCP/IP stack), members cannot be defined to distinguish between the individual server application instances. That is, the combination of the IP address, port, and protocol represents all of the applications sharing the port. Therefore, you cannot selectively quiesce workload requests to only some of the applications sharing the port. Consequently, if you quiesce a member that represents a shareport group, all of the application instances in the group are quiesced.

If an Agent is stopped or fails and is restarted, the quiesce states of any members it might have previously owned are lost. If this occurs, reenter the appropriate quiesce commands to regain the quiesce states that existed during the previous instance of the Agent.

If a member is moved from one system in a sysplex to another using sysplex functions such as VIPA backup, the quiesce state does not move with it. For example, if quiesced member A is using IP address 1.1.1.1 on system X and system X fails, IP address 1.1.1.1 could be moved to system Y. Member A would

no longer be quiesced on the new system and, assuming that the application that member A is assigned to is available and active on system Y, new workload requests would be distributed to this application.

If you plan to take a sysplex system out of service, simply stopping the Agent running on that system is not always a wise alternative to issuing a system-level quiesce command on that system's Agent. If an Agent is simply shut down, there are rare cases where a load balancer might choose to continue routing new workload requests to the applications on that system.

z/OS Load Balancing Advisor configuration example

This topic includes a specific configuration example of the z/OS Load Balancing Advisor, two Load Balancing Agents, and some customization of PROFILE.TCPIP considerations.

In this example, as shown in Figure 148 on page 1188, load balancer LB1 distributes workload requests to two z/OS systems in a sysplex, SYSA and SYSB. SYSA is a CINET configuration with two TCP/IP stacks. SYSB is an INET configuration with one TCP/IP stack. The load balancer is connected to a LAN that also connects to each target TCP/IP stack in the sysplex, including the TCP/IP stack where the Advisor is running. All addresses in this example are IPv4, but the Advisor and Agents are enabled for IPv6.

This example configuration does not use subplexing or AT-TLS.

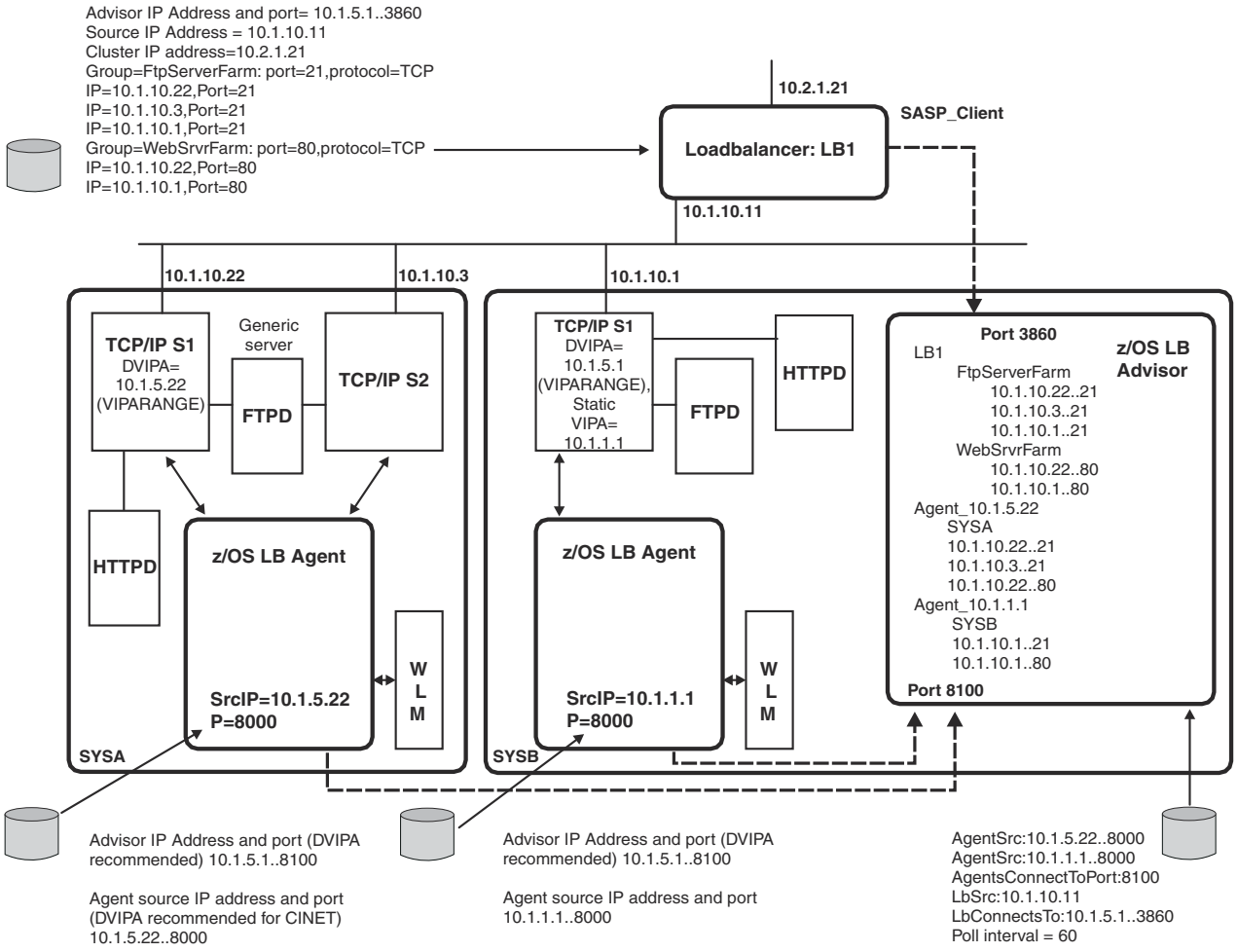


Figure 148. z/OS Load Balancing Advisor configuration example

Load balancer configuration details

The load balancer distributes two types of workload requests in this example, FTP and HTTP traffic, both of which use the TCP protocol. (Keep in mind that UDP workload requests can also be distributed if wanted.) On SYSA, one FTP server is running, which is shared as a generic server between the two

TCP/IP stacks. Also, on SYSA, an HTTP server is bound to stack S1. On SYSB, an FTP and HTTP server are running.

The load balancer must be configured with the IP address and port of the Advisor's load balancing connection socket. In this example, the TCP/IP stack on SYSB has defined 10.1.5.1 as a dynamic VIPA. In addition, this same address and port is defined to the Advisor on the lb_connection_v4 statement in the Advisor's configuration file. The load balancer also uses one of its interfaces to communicate with the Advisor. The IP address assigned to this interface must be coded in the Advisor's lb_id_list statement. In this example, the load balancer uses the interface assigned to the address 10.1.10.11. For information on how to determine which interface the load balancer will use to communicate with the Advisor, consult the specific load balancer documentation.

The load balancer advertises its cluster IP address, 10.2.1.21, so that clients that want to connect to specific applications in the sysplex will connect to this cluster IP address as a proxy. After the connection reaches the load balancer, to determine the actual target of the request, the load balancer consults the information that the Advisor has provided as well as possibly examining the content of the packet. The load balancer might substitute the target IP address in the packet header with the IP address of the member that is best suited to receive the new workload requests, or forward the packet as is using the appropriate MAC address. For example, if a connection request came to the load balancer (10.2.1.21) for port 21 using protocol TCP, the load balancer would forward the packet to either 10.1.10.22, 10.1.10.3, or 10.1.10.1, depending upon which member is the better candidate at that point in time.

To distribute FTP workload requests to the sysplex, a group called FtpServerFarm is defined to the load balancer. The load balancer maps the cluster IP address, 10.2.1.21, port 21, and protocol TCP to this group. In other words, if the load balancer receives a TCP connection with destination 10.2.1.21, port 21, it consults this group to find a member to which it can forward the connection. Within this group are three members that can handle FTP connections, representing target applications within the sysplex. One target can be reached at 10.1.10.22 on port 21, the second at 10.1.10.3 on port 21, and the third at 10.1.10.1 on port 21. The target applications represented by these members do not necessarily all have to be available at all times. The load balancer avoids trying to forward connections to target applications that are not currently available. Therefore, the list of target applications represented by the members in the group should be the entire set of possible members that could handle this workload, now or in the foreseeable future.

To distribute HTTP workload requests to the sysplex, a group called WebSrvrFarm is defined to the load balancer. The load balancer maps the cluster IP address, 10.2.1.21, port 80, and protocol TCP to this group. In other words, if the load balancer receives a TCP connection with destination 10.2.1.21, port 80, it consults this group to find a member to which it can forward the connection. Within this group are two members that can handle HTTP connections, representing target applications within the sysplex. One target can be reached at 10.1.10.22 on port 80, and the second can be reached at 10.1.10.1 on port 80.

Advisor configuration details

There are two aspects of Advisor configuration, the Advisor configuration file itself, and the underlying PROFILE.TCPIP changes that go along with the remainder of the z/OS Load Balancing Advisor system configuration.

The following example is the Advisor configuration file:

```
debug_level 7                       # Error, Warning, Events-- the default

update_interval 60                  # Agents update every  minute-- the default

wlm serverwlm                       # Request server-specific WLM weights

port_list
{
   21    wlm basewlm                    # Use system WLM weights for FTP
}

lb_connection_v4 10.1.5.1..3860   # DVIPA load balancer connects to

lb_id_list
{
   10.1.10.11                       # Load balancer's SASP client interface
```

```
}

agent_connection_port 8100        # Port Agents connect to

agent_id_list
{
   10.1.1.1..8000                 # This system's Agent source
   10.1.5.22..8000                # SYSA's Agent source
}
```

In this example Advisor configuration file, the debug level is set to 7 in the optional debug_level statement. The value of 7 is the default value, so this statement is redundant but is shown for completeness. At the default level of 7, messages are written to the log if they are at error, warning, or event level. Messages at other debug levels, such as info or debug, are not written to the log file.

The optional update_interval statement is set to 60 seconds, which is also the default. This means that each Agent updates the Advisor with new information every minute. For some load balancer implementations, depending upon load balancer configuration, it might also determine how often the load balancer is updated with new information from the Advisor.

The wlm statement specifies the default WLM recommendation type to be used for all groups when calculating the net weights. The value of severwlm indicates that server-specific WLM recommendations will be requested of each Agent, unless overridden by the port_list statement (see next paragraph). Although server-specific WLM is the WLM recommendation type to be used for all groups except FTP in this example, there is still a possibility that WLM system weights might have to be used for some or all groups. For further details on the serverwlm value on the wlm statement, see z/OS Communications Server: IP Configuration Reference.

The port_list statement contains one port number, 21. The wlm keyword indicates that the WLM recommendation type will be overridden for all members using port 21 (FTP), to use WLM system weights rather than server-specific WLM weights when calculating the net weight. Multiple port numbers can appear in the port_list statement on separate lines, if you want to use WLM system weights for other groups of members.

The lb_connection_v4 statement includes DVIPA 10.1.5.1 and port 3860 (the default) as the address and port that load balancers use to connect to the Advisor. The load balancer specifies this address and port when defining the location of the Advisor.

The lb_id_list statement contains the address 10.1.10.11, which is the source IP address of the load balancer when the load balancer connects to the Advisor as a SASP client. If more than one load balancer is used to distribute workload requests to the sysplex, each load balancer needs to be represented in this statement list.

The agent_connection_port statement specifies that port 8100 is used to listen for connections from Agents in the sysplex. This same port appears on each Agent's advisor_id statement. This port is reserved on the TCP/IP stack that the Advisor runs on, and on any TCP/IP stacks that the Advisor could be moved to in the event of failure. Using this port, the Advisor opens a listening socket on the IPv4 or IPv6 unspecified address (0.0.0.0 or ::, respectively), depending upon the TCP/IP stack's IPv6 capability.

The agent_id_list statement contains the source IP address and port of each Agent in the sysplex. The 10.1.1.1 address and the associated port of 8000 represent the source IP address and port that the Agent on SYSB uses to communicate with the Advisor. This same address and port combination appears on the agent_connection statement in the Agent's configuration file on SYSB. The 10.1.5.22 address and the associated port of 8000 represent the source IP address and port that the Agent on SYSA uses to communicate with the Advisor. This same address and port combination appears on the agent_connection statement in the Agent's configuration file on SYSA.

Agent configuration file on SYSB

The following example is the Agent configuration file on SYSB:

```
debug_level 7                  # Error, Warning, Events

advisor_id 10.1.5.1..8100      # DVIPA of Advisor Agent connects to
```

```
host_connection 10.1.1.1..8000  # Source address and port this Agent
                                #   uses to connect to the Advisor
```

In this example Agent configuration file for SYSB, the debug level is set to 7 in the optional debug_level statement. The debug_level statement for an Agent functions similarly to the way it functions for the Advisor.

The advisor_id statement is configured with the address 10.1.5.1 and port 8100. This tells the Agent which address and port the Advisor is using for connections from Agents. The address 10.1.5.1 is configured as a DVIPA on the Advisor's TCP/IP stack. The port of 8100 also appears on the agent_connection_port statement in the Advisor's configuration file, and is also reserved on the Advisor's TCP/IP stack.

The host_connection statement is configured with the address 10.1.1.1 and port 8000. This is the source IP address and port that this Agent uses when connecting to the Advisor. The address is defined as a static VIPA on the TCP/IP stack that the Agent will run on. This address and port must also be specified in the agent_id_list statement in the Advisor's configuration file. The port, 8000, is also reserved in PROFILE.TCPIP of stack S1 on system SYSB.

Agent configuration file on SYSA

The following example is the Agent configuration file on SYSA:

```
debug_level 7                       # Error, Warning, Events

advisor_id 10.1.5.1..8100           # DVIPA of Advisor Agent connects to

host_connection 10.1.5.22..8000     # Source DVIPA and port this Agent
                                    #   uses to connect to the Advisor
```

This Agent configuration file is for the Agent running on SYSA. The debug_level and advisor_id statements are identical to the Agent configuration file on SYSB. The host_connection statement is configured with the address 10.1.5.22 and port 8000. This is the source IP address and port that this Agent uses when connecting to the Advisor. This address and port must also be specified on the agent_id_list statement in the Advisor's configuration file. The port, 8000, is also reserved in PROFILE.TCPIP of stack S1 and S2 on system SYSA. This address is defined as a dynamic VIPA on both TCP/IP stacks on SYSA, in the event that one of the TCP/IP stacks fails.

Customization of PROFILE.TCPIP

Each TCP/IP profile in the sysplex must be updated to accommodate the z/OS Load Balancing Advisor.

The updated portion of PROFILE.TCPIP for stack S1 on system SYSB follows:

```
VIPADYNAMIC
   ;Address LB & Agents use to reach Advisor fall into this subnet
   VIPARANGE DEFINE 255.255.255.0 10.1.5.0
ENDVIPADYNAMIC

DEVICE VIPA41 VIRTUAL 0    ; Static VIPA for Agent's source address
   LINK LVIPA41 VIRTUAL 0 VIPA41
   HOME 10.1.1.1 LVIPA41

PORT
   3860 TCP LBADV          ; SASP Workload Advisor (LB connections)
   8100 TCP LBADV          ; SASP Workload Advisor (Agent connections)
   8000 TCP LBAGENT        ; SASP Workload Agent (Advisor connection)
```

In this example, the address 10.1.5.1 would be within the subnetwork that has been reserved for dynamic VIPAs in the VIPARANGE statement. The load balancer and the Agents use this address to reach the Advisor. Using a dynamic VIPA (DVIPA) facilitates the movement of the Advisor to another TCP/IP stack in the event of failure. This address is defined in the lb_connection_v4 statement in the Advisor's configuration file, in the load balancer as the location of the Advisor [known generically as the SASP Global Workload Manager (GWM)], and on the advisor_id statement in each of the Agent's configuration files.

The address 10.1.1.1 is a static VIPA. The Agent on this system uses this address as its source IP address. Because SYSB is a single stack system (INET), a static VIPA is sufficient. If this were a CINET system like SYSA, a DVIPA would be best. This address appears on the agent_id_list statement in the Advisor's configuration file, as well as on the agent_connection statement in the Agent's configuration file on SYSB.

The ports used for the Advisor and Agent are reserved, as advised. Port 3860 is reserved for the Advisor and is used to communicate with load balancers. This port appears on the lb_connection_v4 statement in the Advisor's configuration file. Port 8100 is also reserved for the Advisor, and is the port that the Agents use to connect to the Advisor. This port appears on the agent_connection_port statement in the Advisor's configuration file, as well as on the advisor_id statement in each of the Agents' configuration files. Port 8000 is reserved for the Agent on this system and is used as the source port for the connection with the Advisor. This port appears on the agent_id_list statement in the Advisor's configuration file, as well as on the agent_connection statement in the Agent's configuration file on this system.

The updated portion of PROFILE.TCPIP for stack S1 on system SYSA follows:

```
VIPADYNAMIC
   ;Address Agent uses as source will fall into this subnet
   VIPARANGE DEFINE 255.255.255.0 10.1.5.0
ENDVIPADYNAMIC

PORT
   3860 TCP LBADV          ; SASP Workload Advisor LB connections,
                           ;  in case Advisor is moved to this stack
   8100 TCP LBADV          ; SASP Workload Advisor Agent connections,
                           ;  in case Advisor is moved to this stack
   8000 TCP LBAGENT        ; SASP Workload Agent Advisor connection
```

In Figure 148 on page 1188, the DVIPA that the Agent uses as a source IP address on this system is shown belonging to stack S1. It could just as easily belong to stack S2, but for the purposes of this example the DVIPA belongs to stack S1.

In this updated portion of PROFILE.TCPIP, the address that the Agent uses as a source address when connecting to the Advisor, 10.1.5.22, is within the subnetwork that has been reserved for dynamic VIPAs on the VIPARANGE statement. Using a dynamic VIPA (DVIPA) facilitates the movement of the Agent to another TCP/IP stack on the same system in the event of failure. This address is defined on the host_connection statement in this Agent's configuration file, and in the agent_id_list statement in the Advisor's configuration file. The DVIPA that the Advisor would use, should the Advisor be moved to this stack, would also fall within this subnetwork.

The port that is used for the Agent is reserved, as advised. Additionally, ports that the Advisor would use if it were to be moved to this TCP/IP stack are also reserved.

The updated portion of PROFILE.TCPIP for stack S2 on system SYSA follows:

```
VIPADYNAMIC
   ;Address Agent uses as source will fall into this subnet
   VIPARANGE DEFINE 255.255.255.0 10.1.5.0
ENDVIPADYNAMIC

PORT
   3860 TCP LBADV          ; SASP Workload Advisor LB connections,
                           ;  in case Advisor is moved to this stack
   8100 TCP LBADV          ; SASP Workload Advisor Agent connections,
                           ;  in case Advisor is moved to this stack
   8000 TCP LBAGENT        ; SASP Workload Agent Advisor connection
```

This updated portion of the TCP/IP profile is identical to that of stack S1 on SYSA. This TCP/IP stack is capable of supporting the Advisor and the Agent running on this z/OS system, should it be necessary to move either to this TCP/IP stack.

Example displays

The following display is from the SYSB Advisor:

```
F LBADV,DISPLAY,LB,INDEX=03
EZD1243I LOAD BALANCER DETAILS 738
LB INDEX      : 03        UUID      : 4C4231
 IPADDR..PORT : 10.1.10.11..50004
 HEALTH       : 7F        FLAGS     : PUSH
 GROUP NAME   : FTPSERVERFARM
  GROUP FLAGS :  BASEWLM
  ProcType    :
    CP : 001  zAAP: 000  zIIP: 000
     IPADDR..PORT: 10.1.10.22..21
      SYSTEM NAME: SYSA      PROTOCOL  : TCP  AVAIL     : YES
      WLM WEIGHT : 00032     CS WEIGHT : 100  NET WEIGHT: 00001
     Raw           CP : 32  zAAP: 00  zIIP: 00
     Proportional  CP : 32  zAAP: 00  zIIP: 00
   FLAGS      :
  IPADDR..PORT: 10.1.10.3..21
   SYSTEM NAME: SYSA      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00032     CS WEIGHT : 100  NET WEIGHT: 00001
     Raw           CP : 32  zAAP: 00  zIIP: 00
     Proportional  CP : 32  zAAP: 00  zIIP: 00
   FLAGS      :
  IPADDR..PORT: 10.1.10.1..21
   SYSTEM NAME: SYSB      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00031     CS WEIGHT : 100  NET WEIGHT: 00001
     Raw           CP : 31  zAAP: 00  zIIP: 00
     Proportional  CP : 31  zAAP: 00  zIIP: 00
   FLAGS      :
 GROUP NAME   : WEBSRVRFARM
  GROUP FLAGS : SERVERWLM
  IPADDR..PORT: 10.1.10.22..80
   SYSTEM NAME: SYSA      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00032     CS WEIGHT : 100  NET WEIGHT: 00001
     Raw           CP : 30  zAAP: 00  zIIP: 44
     Proportional  CP : 10  zAAP: 00  zIIP: 22
   ABNORM     : 00000     HEALTH    : 100
   FLAGS      :
  IPADDR..PORT: 10.1.10.1..80
   SYSTEM NAME: SYSB      PROTOCOL  : TCP  AVAIL     : YES
   WLM WEIGHT : 00031     CS WEIGHT : 100  NET WEIGHT: 00001
     Raw           CP : 30  zAAP: 00  zIIP: 42
     Proportional  CP : 10  zAAP: 00  zIIP: 21
   FLAGS      :
5 OF 5 RECORDS DISPLAYED
```

The following display is from the SYSB Agent:

```
F LBAGENT,DISPLAY,MEMBERS,DETAIL
EZD1245I MEMBER DETAILS 741
LB INDEX      : 03        UUID      : 4C4231
 GROUP NAME   : FTPSERVERFARM
  IPADDR..PORT: 10.1.10.1..21
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : FTPD1     ASID      : 001D RESOURCE  : 0000001A
 GROUP NAME   : WEBSRVRFARM
  IPADDR..PORT: 10.1.10.1..80
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : HTTPD6    ASID      : 0030 RESOURCE  : 00000053
2 OF 2 RECORDS DISPLAYED
```

The following display is from the SYSA Agent:

```
F LBAGENT,DISPLAY,MEMBERS,DETAIL
EZD1245I MEMBER DETAILS 598
LB INDEX      : 03        UUID      : 4C4231
 GROUP NAME   : FTPSERVERFARM
  IPADDR..PORT: 10.1.10.22..21
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : FTPD1     ASID      : 002C RESOURCE  : 0000007D
  IPADDR..PORT: 10.1.10.3..21
   TCPNAME    : S2        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
```

```
   JOBNAME    : FTPD1     ASID      : 002C RESOURCE  : 00000047
 GROUP NAME   : WEBSRVRFARM
  IPADDR..PORT: 10.1.10.22..80
   TCPNAME    : S1        MATCHES   : 001  PROTOCOL  : TCP
   FLAGS      : ANY
   JOBNAME    : HTTPD5    ASID      : 0033 RESOURCE  : 0000005D
3 OF 3 RECORDS DISPLAYED
```

Chapter 22. Automated domain name registration

The automated domain name registration (ADNR) application is a function that dynamically updates name servers with information about sysplex resources in near real time. As resources in the sysplex become available, Domain Name System (DNS) resource records are added to one or more name servers. As those resources become unavailable, the corresponding DNS resource records are removed from the name server. Clients that connect to sysplex resources using DNS names have a greater likelihood of connecting to an available resource in the sysplex. ADNR also removes the administrative burden of manually configuring and updating a name server to represent sysplex resources.

The DNS names managed by ADNR can be application-specific names. All instances of the same application within the sysplex can be represented by the same DNS name. Clients can therefore use one DNS name to connect to any available application instance within the sysplex. ADNR can also manage DNS names that map to specific application instances. These names are used when a client wants affinity to one particular application instance.

In addition to application-specific names, ADNR can also manage DNS names that generically represent the entire sysplex, as well as names that represent individual systems within the sysplex.

How connections are distributed within the sysplex is determined by name server and ADNR configuration. Typically, connections are fairly evenly distributed among the available application instances over time. Connections are not load balanced within the sysplex, as they are with load balancing solutions such as sysplex distributor and the z/OS Load Balancing Advisor.

ADNR supports both IPv4 and IPv6 addresses.

System overview

Figure 149 on page 1195 shows a z/OS sysplex containing four systems, a name server external to the sysplex, a name server in the sysplex, and several clients in the network. The z/OS Load Balancing Advisor and ADNR are running on one of the systems in the sysplex. An instance of the z/OS Load Balancing Agent is active on three of the sysplex systems. Three instances of a server application are active within the sysplex.

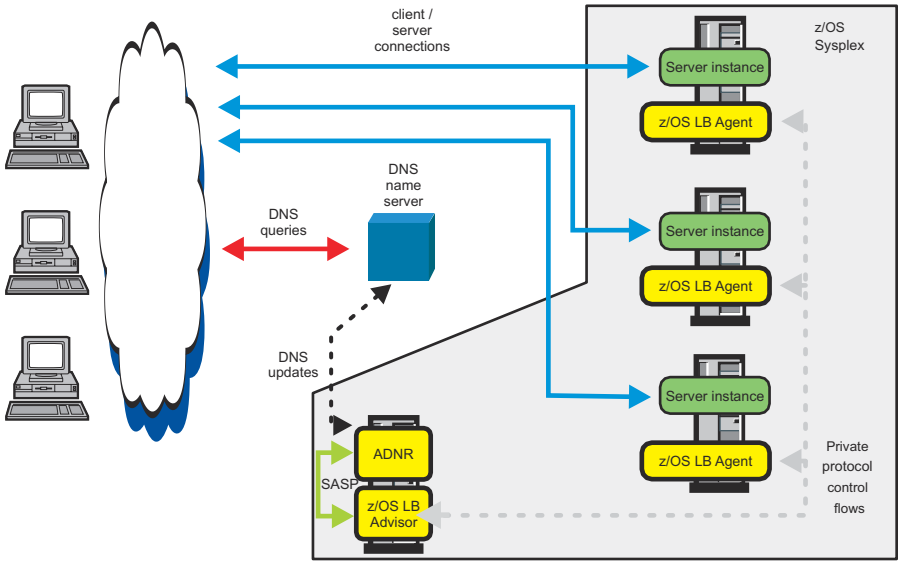


Figure 149. System overview

ADNR is configured with information about sysplex resources to which you want to assign DNS names. These resources are configured on a group basis. For example, to have ADNR manage DNS names in a name server for a sysplex containing several instances of a CICS listener application, all potential

© Copyright IBM Corp. 2000, 2022

CICS listener instances for a given port in the sysplex are defined under the same group in the ADNR configuration file.

ADNR registers its configured information about sysplex resources with a z/OS Load Balancing Advisor application. The Advisor application disseminates this information to the z/OS Load Balancing Agents (Agents), which report back to the Advisor about the availability of the resources registered by ADNR. The Advisor then reports back to ADNR about which of its registered resources are active and which are not. The Advisor subsequently reports to ADNR any changes in the availability of those resources.

For each group of resources that the Advisor reports as available, ADNR adds a DNS name to the name server that represents the entire group of resources in that group, and maps that name to the IP addresses of the available resources in that group. Continuing with the CICS listener example, the IP address of each active CICS listener application is mapped to the name representing the entire group of CICS listener applications. The IP address of each inactive CICS application is not mapped to that DNS name. Clients then connect to their application using the name that ADNR added to the name server. The address or addresses returned to the client's resolver reflect only active application instances. Thus, the client can use one DNS name to connect to any active instance of an application. As application instances become unavailable, the addresses of those unavailable application instances are disassociated from that DNS name in the name server. If all application instances in that group become unavailable, the DNS name representing that group of applications is removed from the name server.

You can also configure ADNR to update the name server with the names of individual server instances, which map to IP addresses that can be used to reach those server instances as those server instances become available. Thus, if a client needs to connect to a specific server instance, that name can be used to make the connection. As each individual server instance becomes unavailable, the DNS name representing the unavailable server instance is removed from the name server.

The name servers that ADNR manages must support RFC 2136, *Dynamic Updates in the Domain Name System (DNS UPDATE)*. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

For administrative purposes, it is advantageous to run ADNR within the same sysplex as the resources it manages. Run ADNR in a sysplex so that it can be moved to provide optimal availability. You can reduce network traffic between ADNR and the Advisor by running ADNR on the same system where the Advisor is running. In addition, the possibility of network outages disrupting communication between ADNR and the Advisor is eliminated by this configuration.

Interaction with name servers

The name servers that ADNR manages require a one-time setup. These name servers must be configured as the primary master name servers for the zones managed by ADNR. The name servers must support RFC 2136, *Dynamic Updates in the Domain Name System (DNS UPDATE)*. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

For more background information about DNS, see “DNS overview” on page 729.

For more information about name server configuration considerations, see “Name server configuration considerations” on page 1210.

Interaction with the z/OS Load Balancing Advisor

The ADNR function is provided by the ADNR application, which uses the z/OS Load Balancing Advisor function.

Requirement: You must configure and deploy the z/OS Load Balancing Advisor and Agent to use ADNR.

An external load balancer is not required.

The ADNR application communicates with the Advisor using the Server/Application State Protocol (SASP). In SASP protocol terms, the Advisor is known as a Global Workload Manager (GWM). The terms Advisor and GWM are used interchangeably. From the GWM's perspective, the ADNR application appears as a load balancer. Despite its appearance as a load balancer to the GWM, ADNR does not perform load balancing. The z/OS Load Balancing Advisor supplies sysplex resource availability and weight information to what

it views as load balancers, including ADNR. ADNR uses the availability information but does not use the weight information.

For more information about this function, see Chapter 21, “z/OS Load Balancing Advisor,” on page 1147.

For more information about z/OS Load Balancing Advisor configuration considerations, see “z/OS Load Balancing Advisor configuration considerations” on page 1208.

Enabling TLS/SSL for ADNR

Consider whether to use AT-TLS for security between ADNR and the z/OS Load Balancing Advisor. AT-TLS provides the ability to authenticate a client, check authorizations, and encrypt data. You must restrict the ability to establish a connection to the Advisor, because sensitive interfaces can be exploited after a connection is accepted by the Load Balancing Advisor. Because ADNR acts as a client to the Load Balancing Advisor SASP port, it must be explicitly authorized to establish its connection to the Load Balancing Advisor.

You can use one or both of the following methods to authorize connection to the z/OS Load Balancing Advisor:

- You can explicitly configure the host_connection_addr keyword on the gwm statement in the ADNR configuration file, and the corresponding lb_id_list statement in the Advisor's configuration file.
- You can establish policies using the z/OS Policy Agent so that ADNR is required to use AT-TLS.

Although the configuration parameters might be sufficient in certain environments where the Load Balancing Advisor and ADNR are inside a secure network (that is, isolated by a firewall and so on), they might not be sufficient in environments in which the network is not considered to be as secure or in which the need to protect against IP address spoofing attacks is important. For more information about using AT-TLS, see Chapter 20, “Application Transparent Transport Layer Security data protection,” on page 1121. For more information about the Advisor, see Chapter 21, “z/OS Load Balancing Advisor,” on page 1147.

Steps for configuring automated domain name registration

The automated domain name registration (ADNR) application dynamically updates name servers with information about sysplex resources. Detail for each step is provided in the corresponding subtopic.

Procedure

Perform the following steps to configure the automated domain name registration (ADNR) application.

1. Decide which sysplex resources should be managed by ADNR.
2. Decide on one or more domain names to be managed by ADNR.
3. Decide which name server or name servers are to be managed by ADNR.
4. Configure the selected name servers to be the primary master name servers for the domain names that ADNR is to manage.
5. Delegate the domain names to be managed by ADNR to the selected name servers from the parent domain's name server.
6. Configure the z/OS Load Balancing Advisor (LBA) function.
7. Define security server profiles for ADNR.
8. Configure ADNR to automatically restart in case of application or system failure. (optional)
9. Configure and start syslogd. (optional, but required to have ADNR write log messages and trace data to syslogd)
10. Configure one ADNR application per sysplex.
11. Customize the TCP/IP profiles of the TCP/IP stacks on which ADNR and the LBA applications are to run. (optional)
12. Start the TCP/IP stacks on which ADNR and the LBA applications are to run.

13. Start the z/OS Load Balancing Advisor and Agent.
14. Start the target applications that are to be managed by ADNR.
15. Start the ADNR application.
16. Verify that the ADNR system is functioning correctly. (optional)

Results

These steps are described in detail in the following subtopics.

Step 1: Decide which sysplex resources should be managed by ADNR

ADNR can manage two types of sysplex resources, including server applications and the traditional DNS mappings of host names to IP addresses. Server applications are represented in ADNR as server groups using the server_groups statement. Traditional DNS host name-to-IP address mappings are represented as host groups using the host_groups statement. For more information about server groups and host groups, see "Identifying the sysplex resources to be managed by ADNR" on page 1204.

ADNR can dynamically create application-specific host names in a name server to represent a cluster of equivalent servers in the sysplex. For example, the name ztelnet.mvsplex.mycorp.com can represent all TN3270E Telnet server applications in the sysplex. Any TCP or UDP server application in the sysplex can be managed by ADNR and have application-specific host names dynamically added and removed from a name server to represent active instances of those servers. As server applications become available, ADNR dynamically adds resource records representing those instances to the name servers managed by ADNR. As those server applications or their systems become unavailable, ADNR dynamically deletes resource records representing those instances from the name servers it manages. These types of DNS names can be used to connect to any active instance of a particular type of server. ADNR also dynamically creates and removes application-specific host names that represent individual instances of server applications as they become available and unavailable.

Most IP addresses in the home lists of sysplex hosts can be dynamically added to the name server as traditional DNS name-to-IP address mappings. As IP addresses are removed from a home list (for example, by issuing a VARY TCPIP,,OBEYFILE command), the DNS resource records representing those IP addresses are dynamically removed from the ADNR-managed name servers. These types of DNS names can be used to connect to any resource in the sysplex, or on a particular sysplex host without regard to which servers are available on that system, such as when using the **ping** or **traceroute** commands.

The IP addresses that are added to DNS by ADNR can be interface addresses, static VIPAs, or dynamic VIPAs (DVIPAs). A small set of addresses cannot be managed by ADNR because of z/OS Load Balancing Advisor restrictions. For more information, see "Step 7: Configure the external load balancers" on page 1168 and the restriction that certain classes of IP addresses must not be coded for members in the load balancer.

You might want to make some sysplex resources visible to some set of clients and not visible to other sets of clients. For example, you might want to add DNS entries to make some sysplex resources visible to intranet clients, but not make the resources visible to Internet clients. There are several ways to accomplish this. Some methods can be accomplished with only name server configuration, others might involve ADNR configuration. For more information, see "Split DNS (views)" on page 1212.

Step 2: Decide on one or more domain names to be managed by ADNR

All resource records added to a name server have a domain suffix related to a name server zone. Typically, an enterprise's domain suffix is something like mycorp.com. Subdomains can exist under a domain, which have the subdomain name added before the parent's domain suffix, such as raleigh.mycorp.com and austin.mycorp.com.

Guideline: Because ADNR should be the only entity updating the zones it manages, ADNR should manage one or more unique sub-zones. For example, if your enterprise has a domain name of mycorp.com, ADNR can be configured to manage resources in a domain called mvsplex.mycorp.com.

The domain suffix of the resource records that ADNR creates is determined by the domain_suffix keyword of the zone keyword of the dns statement in the ADNR configuration file.

Step 3: Decide which name server or name servers are to be managed by ADNR

ADNR can manage name servers that support RFC 2136, *Dynamic Updates in the Domain Name System (DNS UPDATE)*. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

The name servers that ADNR communicates with can be existing name servers, or name servers you set up exclusively for ADNR. Each name server that ADNR is to communicate with is identified on the dns_id keyword of the dns statement.

You might also want to configure one or more secondary name servers for the ADNR-managed zones. Secondary name servers replicate zone data information from the master name server for those zones. Typically, secondary name servers are configured to avoid a single point of failure if a name server fails, and to reduce network traffic by locating secondary name servers in strategic areas of a network so that name server lookups traverse fewer hops in the network.

ADNR does not communicate directly with secondary name servers. Secondary name servers communicate directly with the master name server by performing zone transfers. Ideally, because of the dynamic nature of the data in the zones that ADNR manages, secondary name servers need to be updated as soon as the master name server is updated by ADNR. Otherwise, the secondary name server contains stale information that does not accurately reflect the current availability of sysplex resources. Some name server implementations can minimize the latency with which secondary name servers are updated from their masters, if they have implemented RFC 1996, *A Mechanism for Prompt Notification of Zone Changes (DNS NOTIFY)*. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

Step 4: Configure the selected name servers to be the primary master name servers for the domain names that ADNR is to manage

The name servers that ADNR communicates with must be configured as the primary master name servers of the zones that ADNR will manage. For more information and further references about configuring the zones that ADNR is to manage in the name server, see “Initial zone configuration” on page 1210.

Step 5: Delegate the domain names to be managed by ADNR to the selected name servers from the parent domain's name server

Typically, the DNS domains that ADNR manages are subdomains of an existing DNS domain. For resolvers to reliably find resource records in the ADNR-managed subdomains, the ADNR-managed subdomain must be delegated by its parent domain. This enables resolvers to find resource records in the ADNR-managed subdomain if those resolvers are not pointing directly to the ADNR-managed name server. DNS queries from resolvers can follow the DNS delegation tree downward from the root domain, if necessary, and find an authoritative name server for the ADNR-managed subdomain.

Delegating a subdomain from a parent involves updating the parent domain's zone data file. In the parent's zone data file, an NS record and an associated A or AAAA glue record is added to represent each authoritative name server for the child domain. For example, to delegate the ADNR-managed mvsplex.mycorp.com zone from the mycorp.com zone, the following resource records are added to the zone data file for the mycorp.com zone:

```
mvsplex  86400       IN    NS    mvsnameserver.mvsplex.mycorp.com.
         86400       IN    NS    networknameserver.mvsplex.mycorp.com.
mvsnameserver.mvsplex.mycorp.com.     86400      IN A       10.5.1.1
networknameserver.mvsplex.mycorp.com. 86400      IN AAAA
2001:0DB8:0:0:8:800:200C:417A
```

The example resource records delegate the mvsplex.mycorp.com zone to two authoritative name servers, one of which must be the master name server and the other a secondary name server.

Step 6: Configure the z/OS Load Balancing Advisor function

Configuring and running the z/OS Load Balancing Advisor (LBA) function is a corequisite to implementing ADNR. The z/OS Load Balancing Advisor application communicates with ADNR and serves as ADNR's Global Workload Manager (GWM). The z/OS Load Balancing Agents communicate with the Advisor application and supply it, and ultimately ADNR, with information about the availability of the resources that ADNR has registered with the Advisor. Therefore, each system in the sysplex that contains resources that you want ADNR to manage must be running an Agent, and one system in the sysplex must be running an Advisor.

The Advisor views ADNR as a load balancer, although ADNR does not perform load balancing. ADNR merely uses the information to update the name servers based on the resource availability that the Advisor provides. Therefore, to enable ADNR to connect to the Advisor, the source IP address that ADNR uses to connect to the Advisor must be configured in the Advisor's lb_id_list statement, if AT-TLS is not used. For information about the lb_id_list statement, see z/OS Communications Server: IP Configuration Reference.

For complete information about configuring the Advisor and Agents, see “Steps for configuring the z/OS Load Balancing Advisor” on page 1152. While following those steps, you should skip step 10 to start the TCP/IP stacks that the Advisor and Agents will use until completing the TCP/IP profile customization for those stacks, described in “Step 11: Customize the TCP/IP profiles of the TCP/IP stacks on which ADNR and the LBA applications are to run (optional)” on page 1206. You should also skip steps 13 and 14 in those steps, starting the Agents and Advisor, until you reach “Step 13: Start the z/OS Load Balancing Advisor and Agent” on page 1207.

Step 7: Define security server profiles for ADNR

Create a USERID profile for ADNR as follows:

```
ADDUSER  ADNR     DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(nn) -
         HOME('/RDWR_working_directory') PROGRAM('/bin/sh'))
```

ADNR must have read and write access to the directory specified on the HOME keyword. This directory becomes ADNR's working directory. ADNR creates and deletes temporary files in this directory during its operation. The UID value, *nn*, can be zero or nonzero.

ADNR is a multi-threaded application. If you define an unusually large number of name servers or zones to ADNR, you should check to determine whether the maximum number of threads allowed per process, represented by the THREADSMAX value in BPXPRM*xx*, is going to be exceeded. The number of threads required for ADNR is determined in the following way: (number of dns statements) + (number of zone keywords within all dns statements) + 3. You can customize the maximum number of threads allowed for ADNR by specifying the THREADSMAX keyword on the ADDUSER command.

The program specified on the user ID assigned to run ADNR must be /bin/sh. For more information about specifying a user program and the ADDUSER command, see z/OS Security Server RACF Command Language Reference.

Add ADNR to the STARTED class profile:

```
RDEFINE  STARTED  ADNR.*           STDATA(USER(ADNR))
```

Steps for granting authority to start ADNR

Grant explicit authority to all users that can start ADNR, to prevent unauthorized users from starting it. If you do not grant explicit authority, any user able to issue the START command can start ADNR.

Procedure

Perform the following steps to grant authority to start ADNR.

1. Ensure that the OPERCMDS class is active and RACLISTed, and that RACLIST processing is enabled:

   ```
   SETROPTS CLASSACT(OPERCMDS)
   SETROPTS RACLIST (OPERCMDS)
   ```

2. Define the following OPERCMDS class profile using a security product like RACF:

   ```
   RDEFINE  OPERCMDS (MVS.SERVMGR.ADNR) UACC(NONE)
   ```

3. Grant ADNR access to the OPERCMDS class:

   ```
   PERMIT   MVS.SERVMGR.ADNR CLASS(OPERCMDS) ACCESS(CONTROL) -
                             ID(userid)
   ```

4. Refresh the OPERCMDS class:

   ```
   SETROPTS RACLIST(OPERCMDS) REFRESH
   ```

5. See the EZARACF sample in SEZAINST for specific instructions.

   All commands that you can issue against ADNR are MODIFY commands, with the exception of the STOP command used to stop ADNR. Therefore, you might also want to limit which users are able to issue MODIFY and STOP commands.

Step 8: Configure ADNR to automatically restart in case of application or system failure (optional)

Automatically restarting ADNR minimizes the period of time that name servers do not accurately reflect the status of sysplex resources. This can be accomplished using automation software or by defining an automatic restart manager (ARM) policy. For more information about defining ARM policies, see z/OS MVS Setting Up a Sysplex.

ADNR registers with ARM using the following values:

```
ELEMTYPE=SYSTCPIP
ELEMNAME=EZBADNRelemsuffix
TERMTYPE=ALLTERM
```

The *elemsuffix* value is specified in the arm_element_suffix configuration statement. ADNR registers the element name EZBADNR concatenated with the *elemsuffix* value. If there is no arm_element_suffix statement, the ELEMNAME is EZBADNR. For example, if the following statement is configured:

```
arm_element_suffix SYS1
```

ADNR registers ELEMNAME=EZBADNRSYS1.

When ADNR registers with ARM using these values, then if ADNR fails or the system fails, ADNR can be restarted on any system in the sysplex.

An ARM policy or other automation software should be in place to quickly restart the TCP/IP stack that ADNR is running on if the TCP/IP stack fails. ADNR continues to run after its TCP/IP stack fails, and reconnects to its GWM after the TCP/IP stack recovers.

Although this step is optional, performing it provides high availability to your target applications. In the event that ADNR fails, the name servers it manages are no longer updated with sysplex resource availability information. As a result, client connections might fail because the name server might resolve

their requests to application instances that are not available, or be unable to resolve DNS requests to any application when active application instances might actually exist. When ADNR is restarted, it again accurately reflects the availability of sysplex resources after its convergence period. The *convergence period* is twice the interval at which the Advisor normally updates ADNR with new information. The interval is determined by the update_interval statement in the Advisor's configuration file.

Guidelines:

- Each ADNR instance should have a unique ARM element name within a sysplex.
- Establish an ARM policy with TCP/IP at a lower level than ADNR, so that TCP/IP is restarted before ADNR is restarted. For more information, see z/OS MVS Setting Up a Sysplex.
- To enable ADNR to continue operating on another TCP/IP stack in a Common INET (CINET) environment in the case of TCP/IP stack failure, or to restart on another system in case of system failure, configure ADNR with a unique application-instance DVIPA. The unique application-instance DVIPA is coded on the host_connection_addr keyword of the gwm statement. For more information about unique application-instance DVIPAs, see "Configuring the unique application-instance scenario" on page 373.

Rule: AUTOLOG can be used to start ADNR when the TCP/IP stack is started, but it cannot be used for ADNR recovery. Using AUTOLOG for recovery requires port reservation, but ADNR does not listen on a port. Therefore, ADNR can appear in the TCP/IP AUTOLOG statement, but it cannot appear on the PORT statement. ADNR uses ephemeral ports when connecting to its GWM.

Requirement: ADNR does not run using a system key. Therefore, if you are using ARM registration, the started task ID needs to be permitted with UPDATE authority to the associated IXCARM.SYSTCPIP.EZBADNR* profile in the FACILITY class within the SAF product on your system. To enable ADNR to register to ARM, use the following RACF commands to define the profile and grant update access to the user ID that is assigned to start ADNR:

```
RDEFINE FACILITY IXCARM.SYSTCPIP.EZBADNR* UACC(NONE)
PERMIT IXCARM.SYSTCPIP.EZBADNR* CLASS(FACILITY) ID(ADNR) ACCESS(UPDATE)
SETROPTS REFRESH RACLIST(FACILITY)
```

Restriction: If ADNR is using IPv6 for the GWM connection, or if ADNR is using IPv6 to connect to a name server, movement of ADNR is limited to those TCP/IP stacks that are enabled and configured for IPv6. For considerations for configuring z/OS for IPv6, see z/OS Communications Server: IPv6 Network and Application Design Guide.

Step 9: Configure and start syslogd (optional, but required to have ADNR write log messages and trace data to syslogd)

ADNR writes most log messages and trace data to the syslog daemon (syslogd). A limited number of messages are written to the MVS console, but these are unaffected by syslogd configuration. For ADNR to be able to write log messages and trace data to syslogd, syslogd must be properly configured and started before ADNR is started.

Because you might be running ADNR on the same system as the Advisor, for better readability, you might want to configure syslogd to place Advisor and ADNR log output in separate files. For further information, see "Configuring the syslog daemon" on page 213.

Step 10: Configure one ADNR application per sysplex

ADNR reads configuration data from one file, which can exist as a z/OS UNIX file, a PDS or PDSE member, or a sequential data set. If you plan to enable ADNR to move within the sysplex in the case of failure, the ADNR configuration file or data set should be on shared DASD, to make it accessible to all systems in the sysplex if necessary. The ADNR configuration file is specified on the CONFIG DD statement in the ADNR start procedure. A sample start procedure is provided in EZBADNRS in SEZAINST.

The ADNR configuration file serves five basic purposes:

- Identifies the name servers to update and the zones to be updated in those name servers
- Identifies the GWM to connect to and IP address to bind to for communications with the GWM

- Identifies the sysplex resources to be managed by ADNR
- Uniquely identifies the ADNR instance
- Customizes optional statements

A sample ADNR configuration file is provided in EZBADNRC in SEZAINST.

Tip: Consistent with the Advisor and Agent configuration terminology, any statement or keyword containing the word *connection* refers to a local socket, and any statement containing the word *id* refers to a remote address and possibly a port.

The ADNR configuration file contains statements or keywords that reference other statements or keywords. All such references are made using labels.

Identifying the name servers to update and the zones to be updated in those name servers

Use the dns statement to identify the location of a name server to be managed by ADNR. This statement has a dns_id keyword that contains the IP address of a name server and optionally a port. You can code multiple dns statements.

Rule: In general, for data integrity reasons, multiple dns statements should not refer to the same name server. However, if you are configuring ADNR to work with a split DNS configuration, this is acceptable. For more information, see “Split DNS (views)” on page 1212.

The dns statement also contains one or more zone keywords that identify a zone in the name server to be updated. The domain suffix specified on the domain_suffix parameter must represent a zone previously configured in the name server that is being updated. For more information, see “Initial zone configuration” on page 1210.

The zone keyword contains three optional parameters, update_key, transfer_key, and ttl.

- The update_key parameter and the transfer_key parameter enable the use of digital signatures on requests sent from ADNR to the name server. The digital signatures provide a way for the name server to authenticate ADNR as a client that is authorized to update the name server and to receive zone transfer information. These digital signatures are called transaction signatures (TSIG). Use of TSIGs requires coordination between ADNR configuration and name server configuration. For more information about TSIG security, see “Authorizing dynamic updates” on page 1210 and “Authorizing zone transfers” on page 1211.
- The ttl parameter determines how long resolvers and non-authoritative name servers keep ADNR-managed resource records cached. For more information, see “Near real-time availability information of sysplex resources” on page 1209.

Identifying the GWM to connect to and IP address to bind to for communications with the GWM

Specify the location of the GWM on the gwm_id keyword of the gwm statement. This keyword contains an IP address and optionally a port that the GWM uses to listen for connections from load balancers. From the GWM's point of view, ADNR is considered a load balancer.

Specify the IP address that ADNR will bind to for communications with the GWM on the host_connection_addr keyword of the gwm statement. If you are using AT-TLS, the host_connection_addr keyword is optional.

Rules:

- The IP address specified on the gwm_id keyword of the gwm statement must match the address specified on the lb_connection_v4 statement or the lb_connection_v6 statement in the Advisor configuration file.
- The port specified on the gwm_id keyword of the gwm statement must match the port specified on the Advisor's lb_connection_v4 or lb_connection_v6 statement, depending upon whether you specify an IPv4 or IPv6 address, respectively, on the gwm_id keyword.

- If an IPv4 address is specified on the gwm_id keyword of the gwm statement, an IPv4 address must be specified on the host_connection_addr keyword of the gwm statement. Similarly, if an IPv6 address is specified on the gwm_id keyword of the gwm statement, an IPv6 address must be specified on the host_connection_addr keyword of the gwm statement.

Guideline: For high availability, use a unique application-instance DVIPA for the address on the host_connection_addr keyword. This enables ADNR to be moved to another TCP/IP stack or another system in the event of TCP/IP stack or system failure.

Identifying the sysplex resources to be managed by ADNR

Specify which sysplex resources should be managed by ADNR using the host_group and server_group statements.

Host groups

The host_group statement identifies the set of IP addresses to update in a name server that represents a group of hosts. ADNR updates the name server with the intersection between the IP addresses configured to ADNR in the host_group statement and the IP addresses that exist in the home lists of the hosts in the sysplex. The DNS names that are dynamically added to the name server take the form *host_group_name.domain_suffix*, where *host_group_name* is the ADNR administrator-defined name of the group of hosts being registered to the GWM, and *domain_suffix* is the domain suffix name specified in a zone parameter on a dns statement. To construct the DNS name, the value of the host_group_name keyword of the host_group statement is used, followed by the value of the domain_suffix keyword of the zone referenced in the host_group statement. The intervening dot is supplied by ADNR; do not explicitly code the dot.

ADNR can also update name servers with DNS names representing individual host instances within the sysplex, using the member keyword. ADNR updates name servers with the intersection between the IP addresses configured for the member and the set of IP addresses that exist in the home list of a particular host in the sysplex. The DNS names that are dynamically added to the name server take the form *host_name.domain_suffix*, where *host_name* is the ADNR administrator-defined name of the member being registered to the GWM, and *domain_suffix* is the domain suffix name specified in a zone parameter on a dns statement. To construct the DNS name, the value of the host_name keyword of the host_group statement is used, followed by the value of the domain_suffix keyword of the zone referenced in the host_group statement. The intervening dot is supplied by ADNR; do not explicitly code the dot. If the host_group statement contains a member keyword containing an optional host_name parameter, the name server is updated with these types of DNS names.

A host_group statement can have multiple member keywords within it. One of the coded member keywords does not have to have a host_name parameter. The member without a host_name parameter is sometimes referred to as the *unnamed member*, while members with a host_name parameter are sometimes referred to as *named members*. The IP addresses associated with the unnamed member are ones that can potentially be dynamically added to the name server with the DNS name of the form *host_group_name.domain_suffix*. Configuring an unnamed member is not required. However, a virtual unnamed member is created for you if you do not configure one.

Result: The IP addresses associated with the unnamed member are the union of all IP addresses explicitly configured in the unnamed member (if configured), and all IP addresses configured to all named members in that host group. In contrast, only the IP addresses explicitly configured in a named member are associated with the named member.

Typically, you configure one named member per system in the sysplex. Each named member contains IP addresses that cannot move to other systems in the sysplex.

Guideline: To provide reachability to the sysplex or system in case of interface failure, configure named members of host_group statements with static VIPAs. In addition, you can also code an unnamed member, and associate with it IP addresses that could potentially be moved from one system to another (DVIPAs).

Host names added to a name server using ADNR might be useful for situations when you want connectivity to the sysplex or a host in the sysplex, but the presence of a particular server is not required. For this reason, host group names would be useful for utilities such as ping or traceroute.

Server groups

The server_group statement identifies the set of IP addresses to update in a name server that represents a cluster of equivalent servers in the sysplex. ADNR updates the name server with the intersection between the IP addresses configured to ADNR in the server_group statement and the IP addresses on which each server in the group can be reached in the sysplex. The DNS names dynamically added to the name server take the form *server_group_name.domain_suffix*, where *server_group_name* is the ADNR administrator-defined name of the group of servers being registered to the GWM, and *domain_suffix* is the domain suffix name specified in a zone parameter on a dns statement. To construct the DNS name, the value of the server_group_name keyword of the server_group statement is used, followed by the value of the domain_suffix keyword of the zone referenced in the server_group statement. The intervening dot is supplied by ADNR; do not explicitly code the dot.

ADNR can also update name servers with DNS names representing individual server instances within the sysplex, using the member keyword. ADNR updates name servers with the intersection between the IP addresses configured for the member and the set of IP addresses on which the server can be reached on a particular host in the sysplex. The DNS names dynamically added to the name server take the form *server_name.server_group_name.domain_suffix*, where *server_name* is the ADNR administrator-defined name of the member being registered to the GWM, *server_group_name* is the ADNR administrator-defined name of the group of servers being registered to the GWM, and *domain_suffix* is the domain suffix name specified in a zone parameter on a dns statement. To construct the DNS name, the value of the server_name parameter of the server_group statement is used, followed by the value of the server_name keyword of the server_group statement, followed by the value of the domain_suffix keyword of the zone referenced in the host_group statement. The intervening dot is supplied by ADNR; do not explicitly code the dot. If the server_group statement contains a member keyword containing an optional server_name parameter, the name server is updated with these types of DNS names.

A server_group statement can have multiple member keywords within it. One of the coded member keywords does not have to have a server_name parameter. The member without a server_name parameter is sometimes referred to as the *unnamed member*, while members with a server_name parameter are sometimes referred to as *named members*. The IP addresses associated with the unnamed member are ones that can be dynamically added to the name server with the DNS name of the form *server_group_name.domain_suffix*. Configuring an unnamed member is not required. However, a virtual unnamed member is created for you if one is not configured.

Result: The IP addresses associated with the unnamed member are the union of all IP addresses explicitly configured in the unnamed member (if configured), and all IP addresses configured to all named members in that server group. In contrast, only the IP addresses explicitly configured in a named member are associated with the named member.

Typically, if affinity to a particular server instance is important, you configure one named member per server instance in the sysplex. If affinity to a particular server instance is not important, configuring only an unnamed member suffices.

Guideline: For high availability to the target applications to prevent reachability problems in the event of interface failure, IP addresses in server_group statements should either be static or dynamic VIPAs. If the target application is enabled to move from one TCP/IP stack to another, DVIPAs are appropriate for the server group. Otherwise, static VIPAs are appropriate.

Uniquely identifying this ADNR instance

ADNR appears as a load balancer to the GWM that it connects to. Each load balancer must uniquely identify itself to the GWM. Therefore, each ADNR instance must be uniquely identified, to distinguish it from other ADNR instances, as well as external load balancers that connect to a GWM using SASP. Use the uuid (universally unique ID) statement of the ADNR configuration file to uniquely identify an ADNR instance.

According to the SASP protocol, each load balancer should be universally unique from all other load balancers. In practical terms, only load balancers that connect to the same GWM must be unique.

Guideline: To prevent possible future uuid conflicts among load balancer and ADNR, the uuid configured for each ADNR instance should be universally unique.

Customizing optional statements

Customize the optional statements in the ADNR configuration file.

The optional debug_level statement determines how much log data is captured in the ADNR's log file.

Restriction: In most cases, you should not customize the debug_level statement, unless directed to do so by an IBM Service representative. Adding additional types of trace data can cause the amount of data captured to become voluminous. Reducing the amount of trace data from the default value might make diagnosing a problem more difficult.

Technically, the host_group and server_group statements are optional. They can be omitted to flush a name server of all ADNR-managed resource records. However, if you omit these statements, ADNR will not manage name servers for the sysplex until the ADNR configuration file is updated with host_group and server_group statements and the MODIFY REFRESH command is issued. When no host_group or server_group statements are configured, the ipaddrlist statement is also optional. For more information, see “Flushing a zone” on page 1214.

The key statement defines the key file used in creating digital signatures, which is used when specifying transaction signatures (TSIGs) for zone updates and zone transfers. For more information about keys, see “Authorizing dynamic updates” on page 1210 and “Authorizing zone transfers” on page 1211.

Guideline: Key files must have access permissions that allow ADNR to read the file, either as the file's owner or as a member of the associated group.

Step 11: Customize the TCP/IP profiles of the TCP/IP stacks on which ADNR and the LBA applications are to run (optional)

The TCP/IP profiles can optionally be customized to better accommodate ADNR, the Advisor, and Agents. For information about Advisor and Agent TCP/IP profile customization, see “Step 5: Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run (optional)” on page 1163. You might have already performed the steps for customizing the Advisor and Agent TCP/IP profiles in “Step 6: Configure the z/OS Load Balancing Advisor function” on page 1200.

If you use dynamic VIPAs for ADNR, you need to configure the appropriate TCP/IP profiles in the sysplex for the DVIPA definition and usage. The preferred definitions include VIPARANGE with MOVEABLE NONDISRUPTIVE. For more specific information, see “Using dynamic VIPAs” on page 369.

If you are using AT-TLS, you need to perform several steps to set up the z/OS Policy Agent, enable ADNR's TCP/IP stack for AT-TLS, and so on. For more information, see “Step 5: Customize the TCP/IP profiles of the TCP/IP stacks on which the Advisor and Agents are to run (optional)” on page 1163. For ADNR, you need to customize the AT-TLS policies that are identified with the CommonTTLSConfig and TTLSConfig statements.

Tip: Because ADNR and the Agent are both client applications to the Advisor, you can define AT-TLS policy rules similar to the LBAdvisorAgentRule and the LBAgentRule for ADNR.

Step 12: Start the TCP/IP stacks on which ADNR and the LBA applications are to run

You can start ADNR before or after you start the TCP/IP stack it uses. If the TCP/IP stack that ADNR uses terminates or is not yet started, ADNR remains active and establishes communication with the TCP/IP stack when it becomes active. For additional information about the Advisor and Agents, see “Step 1: Start the TCP/IP stacks that the Advisor and the Agents will use” on page 1171.

Step 13: Start the z/OS Load Balancing Advisor and Agent

You can start the z/OS Load Balancing Advisor and Agents before or after you start ADNR. If you start ADNR before you start the Advisor, an eventual action message, EZD1272E, is displayed until the Advisor is active and a connection is successful. When a successful connection is established, ADNR issues message EZD1270I, and the Advisor issues message EZD1263I indicating that a load balancer connected. At the same time, the eventual action message is converted to a non-action message (DOMed). If Agents were not active at the time that ADNR and the Advisor established a connection, Agents report on any ADNR-registered resources that the Agent owns when the Agent becomes active. ADNR updates its name servers with that information at that time.

Step 14: Start the target applications that are to be managed by ADNR

ADNR can manage TCP and UDP server applications. No modifications are necessary to these applications, their configurations, or start procedures.

Step 15: Start the ADNR application

As ADNR connects to the Advisor, an MVS console message appears on the MVS console of the system on which ADNR is running. Simultaneously, a message appears on the Advisor's MVS console, indicating that a load balancer connected. If ADNR fails to connect to the Advisor, an eventual action message is displayed on ADNR's MVS console until the connection is successful. ADNR must be started from a job control procedure that is in a procedure library. It cannot be started under BPXBATCH. The IBM-supplied program properties table has entries to make ADNR run non-swappable. You should not override this entry to make ADNR run swappable.

Step 16: Verify that the ADNR system is functioning correctly (optional)

View the MVS console of the ADNR system after you have started it to verify that the application started correctly and is still running. If there are any failure messages, see the appropriate message description for the appropriate corrective action. View the syslogd file of the ADNR system to see whether any error or warning messages were issued.

Use the following commands to verify that ADNR is functioning correctly:

- If connectivity cannot be established, ADNR issues the eventual action message `EZD1272E, ADNR CONNECTION ATTEMPT TO GWM gwmipaddress FAILED`. If this message is issued and the GWM is active, verify that routing exists between the IP addresses used by the GWM and ADNR.
- Verify that ADNR is started and connected to the expected Advisor by issuing ADNR's MODIFY *procname*,DISPLAY,GWM,DETAIL command. Verify that the GWM IP address and port is the expected IP address and port of the Advisor to which you intended to connect.
- From the Advisor, issue the MODIFY *procname*,DISPLAY,LB command to list the LB indexes known to the Advisor. Identify the LB index representing ADNR, and then issue the Advisor MODIFY *procname*,LB,INDEX=*lb-index* command. Verify that each member you have registered that you expect to be available in the sysplex displays as available in the AVAIL field. If there is a discrepancy in the availability status, check the FLAGS field for possible reasons why a member is not available. For more information about the flags in the FLAGS field for the Advisor, see z/OS Communications Server: IP System Administrator's Commands.
- Verify that server and host groups have been properly registered to the GWM by issuing ADNR's MODIFY *procname*,DISPLAY,GWM,GROUPS,DETAIL command. Verify that each member you have registered that you expect to be available in the sysplex displays as available in the AVAIL field. If there is a discrepancy in the availability status, check the FLAGS field for possible reasons why a member is not available. For more information about the flags in the FLAGS field for ADNR, see z/OS Communications Server: IP System Administrator's Commands.
- Verify that Agents are aware of members that they own by issuing the MODIFY *procname*,DISPLAY,MEMBERS Agent command. Any ADNR-registered IP addresses owned by the system where the Agent is running should appear in the display.

- Verify that ADNR is able to communicate properly with the name servers it manages by issuing ADNR's MODIFY *procname*,DISPLAY,DNS,ZONES,SUMMARY command. Verify that the ZONE STATUS field is or changes to SYNCHRONIZED after the convergence period expires. For the definition of the convergence period, see "Step 8: Configure ADNR to automatically restart in case of application or system failure (optional)" on page 1201. If any zone status begins with the value NOT_RESPONSIVE, there is either a connectivity problem between ADNR and the name server, or there is a configuration mismatch between ADNR and the name server. For more information about diagnosing unresponsive zones, see z/OS Communications Server: IP Diagnosis Guide. If ADNR is unable to properly communicate with a name server, ADNR issues eventual action message EZD1278E, which remains until the problem is resolved.
- Verify that name servers reflect the availability status of the sysplex resources that you have configured with ADNR. Issue ADNR's MODIFY *procname*,DISPLAY,ZONES,ZONEID=*zone_label*,DETAIL command for each zone configured to ADNR, or issue the MODIFY *procname*,DISPLAY,ZONES,DETAIL to view all zones. Verify that the DNS RR STATUS field for each resource record display is PRESENT for each resource that is available in the sysplex.
- Use the z/OS UNIX **dig** command to perform a zone transfer of the ADNR-managed zones, and verify that the zone contents match the expected content, based on the members configured to ADNR and the availability status of those members in the sysplex. The **dig** command to use has the following form:

  ```
  dig @name_server_address  domain_name axfr
  ```

  Tip: If zone transfers were restricted by name server configuration to clients from specific IP addresses or from clients that have digitally signed their requests (TSIG), you might have to issue the **dig** command with the -b option or the -k option to, respectively, force the request to originate from a specific IP address or be digitally signed with the appropriate signature. For more information about the **dig** command, see z/OS Communications Server: IP System Administrator's Commands.
- Verify that ADNR is functioning correctly when using AT-TLS:
  - Use the **pasearch** command from the z/OS UNIX shell to query information from the Policy Agent. For more information about displaying policy based networking information, see z/OS Communications Server: IP System Administrator's Commands.
  - Use the Netstat TTLS/-x command to display z/OS Load Balancing Advisor and ADNR AT-TLS policies. For more information about the Netstat TTLS/-x report, see z/OS Communications Server: IP System Administrator's Commands.

z/OS Load Balancing Advisor configuration considerations

The z/OS Load Balancing Advisor solution consists of a z/OS Load Balancing Advisor application, and one or more z/OS Load Balancing Agents, per sysplex. For more information about the z/OS Load Balancing Advisor solution, see Chapter 21, "z/OS Load Balancing Advisor," on page 1147.

Connectivity considerations

ADNR connects to the Advisor. The Advisor can use an access control list (ACL) to determine which load balancers are permitted to connect to it. From the Advisor's perspective, ADNR is considered to be a load balancer.

Rules:

- If you are not using AT-TLS, the Advisor's lb_id_list statement must include the IP address coded on the host_connection_addr keyword of the ADNR gwm statement. If you are using AT-TLS for all ADNR-Advisor connections, the lb_id_list statement and host_connection_addr keyword are ignored. If you are defining external load balancers that are not using TLS/SSL, the lb_id_list statement is required.
- The IP address and port coded on the Advisor's lb_connection_v4 or lb_connection_v6 statement must match the IP address and port coded on the gwm_id keyword of the ADNR gwm statement, depending on the address family used.

Guideline: For availability reasons, the IP address coded on the gwm_id keyword of the ADNR gwm statement should be a unique application-instance DVIPA.

Near real-time availability information of sysplex resources

To take advantage of the near real-time availability information that ADNR can provide regarding sysplex resources, clients must query the ADNR-managed name servers relatively frequently rather than use cached information from a previous name server query. The length of time that name server information is cached is typically controlled by the Time To Live (TTL) value received from the name server. A TTL is associated with each resource record in a name server, and the TTL information is part of the information that the resolver receives when the name server successfully responds to a query. ADNR provides a mechanism to define the TTL for name server resources it manages.

ADNR permits the TTL to be defined to a specific value per zone, or it can be set to a value determined by the GWM, which is the default. When the TTL value is the default value determined by the GWM, it takes on the update_interval value in the Advisor's configuration file.

Guideline: Allowing the GWM to set the TTL for an ADNR zone by using the default TTL value enables clients to obtain the most accurate, near real-time sysplex resource availability information, without wasting network resources with unnecessary DNS queries.

Because ADNR can be configured to use a GWM-defined interval to set the TTL value, the GWM can dynamically change this interval. The z/OS Load Balancing Advisor does not support dynamically changing this interval, but a future Advisor implementation or other future GWM implementations might. If this occurs, ADNR has the capability to find out about the new GWM interval only when the connection to the GWM is re-established.

z/OS Load Balancing Advisor and Agent operational considerations

Several operational characteristics of the z/OS Load Balancing Advisor and Agent can affect ADNR.

Advisor operational considerations

For ADNR to accurately maintain the resource records that represent available sysplex resources in its managed name servers, the Advisor and all Agents in the sysplex need to be active. If the Advisor stops, ADNR can no longer receive information about the current availability of sysplex resources, and consequently, the resource records in the name servers that ADNR manages eventually become stale. That is, resource records are left in the name server representing sysplex resources that are no longer available, or resource records are not added to the name server to represent sysplex resources that have become available. Because of the ramifications of the Advisor becoming unavailable, you should perform the necessary configuration for optimal availability of the Advisor. For more information, see “Step 1: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)” on page 1153.

Agent operational considerations

If an Agent in the sysplex stops, all resource records in the ADNR-managed name servers that represent sysplex resources managed by that Agent are removed from the name server. This makes it appear as though the resources are unavailable when they actually might be available. For this reason, it is important to perform the necessary configuration for optimal availability of each Agent. For more information, see “Step 1: Configure the Advisor and Agents to automatically restart in case of application or system failure (optional)” on page 1153.

The Agent is also capable of making resources appear unavailable to ADNR by quiescing them with the Agent's MODIFY QUIESCE command. All resource records in a name server associated with a quiesced member are removed and are not added back until the member is enabled with the Agent's MODIFY ENABLE command, or if the Agent is stopped and restarted.

Name server configuration considerations

The name servers that ADNR manages must support RFC 2136, *Dynamic Updates in the Domain Name System (DNS UPDATE)*. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

Many of the subjects discussed in this topic are applicable to other name server implementations. For BIND 9 name servers, specific instructions are provided for how to accomplish configuration objectives. For other name server implementations, consult the appropriate product documentation.

Initial zone configuration

ADNR cannot create a zone in a name server. It can only add and delete resource records from an existing zone. The zones that ADNR manages must initially be configured to the appropriate name servers. This involves defining the zone and a small set of resource records in the zone. This includes the SOA record, the appropriate NS records, and the A or AAAA glue records.

Rule: Do not configure any valuable A or AAAA resource records in the zone data files that ADNR will manage. ADNR deletes any type A or AAAA resource records that were not added by ADNR, with the exception of glue resource records (see “Updates to an ADNR-managed zone” on page 1210). These resource records are deleted when ADNR starts, or during MODIFY REFRESH processing.

Restriction: ADNR does not support DNSSEC signed zones.

Authorizing dynamic updates

Name servers can be configured to allow dynamic updates from only specific entities. If a name server that ADNR updates is configured to restrict which entities can update the name server, ADNR must be specifically permitted to update that name server. Name servers typically permit dynamic updates from a predetermined set of source IP addresses, sometimes referred to as an access control list (ACL), or they might require authentication with digital signatures, sometimes referred to as transaction signatures (TSIG). Authentication using digital signatures is much more secure than authenticating by source IP address, because the latter is subject to address spoofing. Furthermore, the source IP address that ADNR uses might not be entirely predictable, unless deliberate steps are taken in the TCP/IP profile to make it predictable through mechanisms like job-specific source IP address specification or other forms of SOURCEVIPA configuration. Other possible options for lessening the impact of the unpredictability of the source IP address that ADNR uses include using a subnet in the name server's ACL. However, this allows updates from any entity in that subnet, which compromises security.

Guideline: Digital signatures (TSIG authentication) provide more robust authentication than source IP address permissions (ACL).

For the BIND 9 name server, dynamic updates are allowed by ACL using the allow-update statement, or by digital signatures using the update-policy or the allow-update statement.

When dynamic updates are permitted using digital signatures (TSIG), the name server and ADNR must be configured with the same shared cryptographic key. You can generate the key using the dnssec-keygen utility. Then you define the key to the name server and reference the key from name server zone definitions that want to use it.

You must also define the TSIG key to ADNR using the key configuration statement, and then reference the TSIG key from the update_key keyword of the zone keyword of the dns statement. The key file should be protected from unauthorized access. ADNR must have read access to the file. Both the .key and the .private key files generated by the dnssec-keygen utility must be present for ADNR to properly communicate with the name server, even though only the .key key file name is actually specified on the update_key keyword.

Updates to an ADNR-managed zone

Typically, zones managed by ADNR should not be updated by any entity other than ADNR. This includes manual editing of the zone data file, or dynamic updates from any other source. Many name server

implementations strongly advise against manually editing the zone data file of a dynamic zone, such as the zones managed by ADNR. Editing these files has the potential to corrupt the file, or the updates can be lost. However, in general from a name server perspective, dynamic updates using a tool such as nsupdate are allowed for dynamic zones. In the case of ADNR-managed zones, certain types of dynamic updates can be made by entities other than ADNR. Table 65 on page 1211 shows which types of resource records can be dynamically added and deleted by entities other than ADNR.

Table 65. Dynamic updates of ADNR-managed zones by other entities

Allowed dynamic updates	**Disallowed dynamic updates**
All resource record types other than A and AAAA	A and AAAA resource records that are not considered glue records
A and AAAA resource record types that are considered glue records	
Tip: Glue records are A or AAAA records that store the IP addresses of authoritative name servers.	

Any disallowed resource records that appear in dynamic update requests to ADNR-managed zones are accepted by the name server, but are subject to being deleted by ADNR at a later time.

Rule: ADNR uses a special domain name for determining whether a name server is active and configured correctly to allow dynamic updates. At times, ADNR dynamically performs an update-add and an update-delete of the following resource record:

```
ibmsaspddnsdnsprobe IN A 0.0.0.0
```

In the unlikely event that this same resource record is added to an ADNR-managed zone by an entity other than ADNR, it is subject to being deleted by ADNR in the future.

Update forwarding

Update forwarding is a DNS concept in some name server implementations that enables dynamic updates to be sent to a name server that is not the primary master name server for the zone being updated. The name server that receives the dynamic update request forwards the update to the primary master name server.

Rule: To avoid possible zone data-integrity issues, ADNR must be configured to update only the primary master name server for a zone. That is, the IP address of a name server on the dns_id keyword of the dns statement must be an address of the primary master name server for the zone identified on the domain_suffix keyword, and must not be the IP address of a secondary name server for that zone.

Authorizing zone transfers

Zone transfers are operations that are typically performed between authoritative name servers of a particular zone. A zone transfer occurs when all or part of the contents of a zone is sent from the name server to the requester. ADNR requests zone transfers from the primary master name server of the zones it manages. Name servers can be configured to allow zone transfers only from specific entities. If the name servers that ADNR is to update are configured to restrict which entities can perform zone transfers, ADNR must be specifically permitted to perform them. Name servers typically permit zone transfers from a predetermined set of source IP addresses, sometimes referred to as an access control list (ACL), or they can require authentication using digital signatures, sometimes referred to as transaction signatures (TSIG). Authentication using digital signatures is much more secure than authenticating by source IP address, because the latter is subject to address spoofing. Furthermore, the source IP address that ADNR uses might not be entirely predictable, unless deliberate steps are taken in the TCP/IP profile to make it predictable through mechanisms like job-specific source IP address specification or other forms of SOURCEVIPA configuration.

Guideline: Digital signatures (TSIG authentication) provide more robust authentication than source IP address permissions (ACL).

For the BIND 9 name server, zone transfers are allowed by ACL or by digital signatures, both by using the allow-transfer statement.

You must define the TSIG key to ADNR using the key configuration statement, and then reference the key from the transfer_key keyword of the zone keyword of the dns statement. The key file should be protected from unauthorized access. ADNR must have read access to the file. Both the .key and the .private key files generated by the dnssec-keygen utility must be present for ADNR to properly communicate with the name server, even though only the .key key file name is actually specified on the transfer_key keyword.

Limiting the duration of an outbound zone transfer

With some name server implementations, you can limit the amount of time that an outbound zone transfer is allowed to run. Setting this time interval to a value that is too short can have adverse affects on ADNR's ability to communicate with the name server. ADNR must request full zone transfers from the name servers it manages. If the period of time allowed for a zone transfer is shorter than the amount of time it normally takes to transfer an ADNR-managed zone, ADNR cannot successfully manage that zone; adjust the name server configuration to provide adequate time for a full zone transfer.

For the BIND 9 name server, the amount of time to allow for an outbound zone transfer is specified on the max-transfer-time-out option.

Limiting the total number of simultaneous outbound zone transfers

With some name server implementations, you can limit the total number of simultaneous zone transfer requests that are served. Setting this number too low can have adverse affects on ADNR's ability to communicate with the name server. When setting this limit, take into account the number of secondary name servers that use this name server as the master name server for their zones, as well as the number of ADNR-managed zones that are configured under this name server. During ADNR initialization and after a MODIFY REFRESH command, ADNR requests simultaneous zone transfer requests for each zone it manages. Consider the number of zones configured under the same dns statement when you are setting zone transfer limits in the name server.

For the BIND 9 name server, the total number of simultaneous outbound zone transfers allowed is specified on the transfers-out option.

The .digrc file

ADNR uses the z/OS UNIX **dig** command to perform zone transfers from managed zones. The z/OS UNIX .digrc file can provide defaults for each user for the **dig** command. This file is read and used if it appears in the user's home directory. Because this file can cause ADNR to function incorrectly, ADNR fails initialization if this file is found in ADNR's working directory. However, ADNR does not search for this file after initialization. ADNR's working directory is specified when ADNR is defined to the security server (RACF).

Rule: Do not put the .digrc file in ADNR's home directory.

Split DNS (views)

Split DNS is a method in some name server implementations that enables zone data to appear differently depending on which client queries the name server. For example, Internet clients might be restricted to a subset of resource records in a zone, while intranet clients might have access to all resource records in a zone.

The BIND 9 name server implementation supports split DNS through the view statement.

If you want to configure split DNS for an ADNR-managed zone, there are some special considerations. When implementing split DNS for a domain, the zone data is kept in separate files in the name server for each view of the zone. This implies that the data is disjointed, and in many respects ADNR must treat each view like a separate zone. When performing dynamic updates to a split DNS zone, the name server might determine which view to update based upon the source IP address of the nsupdate client. The BIND 9 name server uses this criteria to determine which view to update, and ADNR is the update client

in this regard. Therefore, ADNR must use separate source IP addresses to update each view. This can be accomplished in the following ways:

- Configure separate ADNR applications for each view.

  By configuring an ADNR application for each view, a unique source IP address can be associated with each ADNR instance using job-specific source IP addressing. The source IP address assigned to ADNR needs to be configured appropriately in the name server. For the BIND 9 name server, the source IP address of the ADNR instance that you want to have update a specific view would appear in that view's match-clients statement.

  The easiest way for each ADNR instance to use unique source IP addresses is to run each instance on separate systems or TCP/IP stacks. If each ADNR instance must run on the same TCP/IP stack, you can use job-specific source IP addressing to provide unique source IP addresses. This involves mapping the ADNR job names to the source IP address to be used using the JOBNAME entry in the SRCIP statement block of the TCP/IP profile. However, each ADNR instance creates multiple address spaces, whose job names are the ADNR start procedure name with a numerical digit appended (for example, the *ADNR* job creates the *ADNR1* and *ADNR2* jobs). To use this method, job-specific source IP addressing has to specify the ADNR start procedure name followed by a wildcard when specifying the job name. Also, to avoid conflicting job names of the spawned address spaces among the multiple ADNR instances, you need to name the start procedures of the ADNR jobs so that the spawned job names do not conflict. For example, the start procedure names of *ADNRP* and *ADNRT* create unique job names among their instances, such as *ADNRP1* and *ADNRT1*; create unique source IP addresses by coding JOBNAME ANDRP* *source_ip_address1* and JOBNAME ADNRT* *source_ip_address2*.

- Understand and exploit TCP/IP's source IP address selection algorithm.

  By understanding and exploiting TCP/IP's source IP address selection algorithm you can configure ADNR and locate name servers in the network, so that ADNR uses appropriate source IP addresses to dynamically update the appropriate views. For more information, see “Source IP address selection” on page 244.

  This technique involves coding separate dns statements in the ADNR configuration file, specifying different IP addresses for each dns statement (of the same multi-homed name server) and using the separate dns statements to update different views. If configured correctly, and depending on the IP address of the name server, the source IP address of ADNR can be influenced to match the IP address in the match-clients statement of the appropriate view in the name server.

  The most predictable way to assign the correct source IP address for each separate view in this scenario is to use the DESTINATION entry in the SRCIP statement block of the TCP/IP profile. For example, if two dns statements with different IP addresses are coded in the ADNR configuration file, where each represents different views in the same name server, two DESTINATION entries are coded in the SRCIP statement of the TCP/IP profile. One DESTINATION entry contains the IP address from one of the dns statements as the destination IP address, and the second DESTINATION entry contains the IP address from the other dns statement as the destination IP address. The source IP address contained in each DESTINATION entry matches one of the IP addresses in the match-clients statement in the name server configuration file of the appropriate view to be updated.

Zone transfer formats

Some name server implementations have the capability to send multiple resource records in a DNS message during a zone transfer operation, instead of just one resource record per DNS message. The process of sending multiple resource records per DNS message is described in RFC 2671, *Extension Mechanisms for DNS (EDNS0)*. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

Sending multiple resource records per DNS message is more efficient than sending just one. ADNR performs zone transfer requests from the name servers it manages. ADNR supports receiving zone transfers in either format, but is more efficient at processing the format that contains multiple resource records per DNS message.

The BIND 9 name server sends multiple resource records per DNS message by default. However, some name server implementations do not support receiving zone transfers using this format. For this reason, some name servers might be configured to use the less efficient format when sending zone data to a secondary name server. The BIND 9 name server supports specification of the zone transfer format on the transfer-format option. This option can be specified on a server statement, which can be used to set the option only when sending zone transfer information to ADNR, so that other name servers are unaffected by the setting.

ADNR configuration considerations

This topic describes ways to change the ADNR configuration file and how to maintain zone data integrity.

Changing the ADNR configuration file

The ADNR MODIFY *procname*,REFRESH command provides a way to dynamically change the ADNR configuration file. Because the command does not support a file name or dataset specification, changes must be made to the same configuration file that was used when ADNR was started.

Restriction: Make all changes to the ADNR configuration file dynamically while ADNR is running. Changing the configuration file without using the MODIFY REFRESH command can create orphaned resource records in a name server. Orphaned resource records are resource records that are left in a name server and that ADNR has lost the ability to manage. The presence of these orphaned resource records might inaccurately direct client connections to sysplex resources that are not actually available.

Tip: Before making changes to the ADNR configuration file, save a backup copy of the original file under a different file name or dataset name.

You can make changes to the ADNR configuration file while ADNR is stopped, instead of using the MODIFY REFRESH command, if you take precautions to prevent orphaned resource records or you are sure that clients will no longer have their resolvers pointing to a name server with orphaned resource records. Unless you use the MODIFY REFRESH command to make changes, the following types of changes to the ADNR configuration file create orphaned resource records in name servers:

- Removing a dns statement
- Removing a zone keyword from a dns statement
- Changing the domain_suffix keyword value on a dns statement
- Changing the IP address or port of a name server on the dns_id keyword of a dns statement

Flushing a zone

Flushing a zone refers to deleting all ADNR-managed resource records from an ADNR-managed zone. Flushing a zone is not a normal occurrence, but might be a useful administrative tool in certain situations. You can flush all zones, or just specific zones.

Flush all zones by removing all host_group and server_group statements from the ADNR configuration file and issuing a MODIFY REFRESH command. You might flush all zones when resource records are inadvertently orphaned in a name server. The orphans can be removed by ADNR by restoring a copy of the previous configuration file, removing all host_group and server_group statements, and then issuing a MODIFY REFRESH command. When the command has completed, you can issue another MODIFY REFRESH command using the latest ADNR configuration file to return to the configuration that you want to use.

You flush individual zones in much the same way as you flush all zones. Instead of removing all host_group and server_group statements from the ADNR configuration file, you remove only those statements that reference the zone or zones that you want to flush.

Maintaining zone data integrity

To ensure the integrity of resource records in the name server zones that ADNR is managing, several guidelines should be followed.

Guidelines:

- Use the MODIFY REFRESH command to change the ADNR configuration.
- If you have modified the ADNR configuration, issue a MODIFY REFRESH command before stopping ADNR. Do not stop and restart ADNR before issuing this command.
- Do not stop ADNR until a MODIFY REFRESH command has completed.
- If message EZD1273I is issued indicating that ADNR was unable to delete a zone during a MODIFY REFRESH operation, take other steps to delete the resource records added by ADNR. This might include flushing the zone, after the underlying problem that prevented ADNR from deleting the zone is resolved. For more information, see “Flushing a zone” on page 1214.
- Do not configure multiple zones for the same domain suffix that references the same multi-homed name server.
- Do not manually edit the name server's zone data file for any zone managed by ADNR.
- Do not allow any entity other than ADNR to perform dynamic updates to the zones managed by ADNR.
- Do not employ update forwarding for ADNR-managed zones.

Steps for using the ADNR application in a sysplex subplexing environment

In a sysplex subplexing environment, there are some additional considerations for using ADNR. These steps are described in detail in the following subtopics.

Before you begin

Each subplex containing resources that were managed by ADNR before the sysplex was split into subplexes requires its own set of z/OS Load Balancing Advisor and Agent applications, and its own ADNR application, to continue to enable ADNR to manage DNS names for those subplex resources. Each subplex ADNR instance will communicate with the Advisor application in its subplex. Each subplex ADNR instance will update separate DNS zones. The ADNR instances for different subplexes cannot update the same zone, because each zone can be updated by only one ADNR application and ADNR can communicate with only one Advisor instance. There must be a one-to-one correspondence between a subplex and a DNS zone.

Procedure

Perform the following steps to use ADNR in a sysplex subplexing environment.

1. Plan how the new subdomains representing each subplex will fit into your DNS hierarchy.
2. Configure the name servers that will be updated for the new subplex domains.
3. Define and configure one Advisor per subplex.
4. Update the Agent configuration files to communicate with the Advisor running in its subplex.
5. Define one ADNR application per subplex.
6. Assign the host_group and server_group statements from the sysplex ADNR configuration to their correct subplex domains.
7. Configure the new ADNR instances to update the name server and zone for its subplex.
8. Configure the new ADNR instances to communicate with the subplex Advisor.
9. (Optional) Update resolver configuration files.
10. Start the TCP/IP stacks, Advisor, Agent, ADNR, and target applications that ADNR will manage.
11. Verify that each subplex ADNR is functioning correctly.

Step 1: Plan how the new subdomains representing each subplex will fit into your DNS hierarchy

Decide whether the new subdomain names should exist at the same level as the existing subdomain representing the sysplex that is being divided into subplexes, whether the new subdomains should become child domains of the existing sysplex domain, or whether there will be a mixture of the two.

For example, if the existing subdomain for the sysplex is mvsplex.mycorp.com, the new subdomains that represent the subplexes can be at the same level in the DNS hierarchy, such as mvsplex1.mycorp.com and mvsplex2.mycorp.com. Alternatively, you can create the new subdomains as child domains of the existing sysplex domain, such as subplex1.mvsplex.mycorp.com and subplex2.mvsplex.mycorp.com.

For the following examples, assume that the sysplex and ADNR are configured as shown in Table 66 on page 1216.

Table 66. Base sysplex and ADNR configuration

Sysplex A Domain=mvsplex.mycorp.com	
Host1	**Host2**
z/OS Load Balancing Agent	z/OS Load Balancing Agent
z/OS Load Balancing Advisor for sysplex A	
ADNR for sysplex A updating zone mvsplex.mycorp.com	

To convert this sysplex to a subplexing environment, one of the subplexes can use the existing domain name as the existing sysplex, and the other subplexes become subdomains of that domain, as shown in Table 67 on page 1216.

Table 67. ADNR application in a sysplex subplexing environment; Example 1

Sysplex A (No DNS domain)	
Host1 Member of subplex EZBT0211 Domain=mvsplex.mycorp.com	**Host2 Member of subplex EZBT0212 Domain=subplex0212.mvsplex.mycorp.com**
z/OS Load Balancing Agent for subplex EZBT0211	z/OS Load Balancing Agent for subplex EZBT0212
z/OS Load Balancing Advisor for subplex EZBT0211	z/OS Load Balancing Advisor for subplex EZBT0212
ADNR for subplex EZBT0211 updating zone mvsplex.mycorp.com	ADNR for subplex EZBT0212 updating zone subplex0212.mvsplex.mycorp.com

Another possibility is to assign each subplex as a child domain of the original sysplex domain, as shown in Table 68 on page 1217. In this case, the existing sysplex domain no longer contains any resource records that represent hosts or server applications, but instead is merely the parent of the new subplex subdomains.

Table 68. ADNR application in a sysplex subplexing environment; Example 2	
Sysplex A Domain=mvsplex.mycorp.com (Only delegates the child domains. Contains no sysplex resources.)	
Host1 Member of subplex EZBT0211 Domain=subplex0211.mvsplex.mycorp.com	**Host2 Member of subplex EZBT0212 Domain=subplex0212.mvsplex.mycorp.com**
z/OS Load Balancing Agent for subplex EZBT0211	z/OS Load Balancing Agent for subplex EZBT0212
z/OS Load Balancing Advisor for subplex EZBT0211	z/OS Load Balancing Advisor for subplex EZBT0212
ADNR for subplex EZBT0211 updating zone subplex0211.mvsplex.mycorp.com	ADNR for subplex EZBT0212 updating zone subplex0212.mvsplex.mycorp.com

Another alternative is to create new domain names for each of the subplexes at the same level of the DNS hierarchy as the existing sysplex domain, and retire the existing sysplex domain name, as shown in Table 69 on page 1217.

Table 69. ADNR application in a sysplex subplexing environment; Example 3	
Sysplex A (No DNS domain)	
Host1 Member of subplex EZBT0211 Domain=mvsplex0211.mycorp.com	**Host2 Member of subplex EZBT0212 Domain=mvsplex0212.mycorp.com**
z/OS Load Balancing Agent for subplex EZBT0211	z/OS Load Balancing Agent for subplex EZBT0212
z/OS Load Balancing Advisor for subplex EZBT0211	z/OS Load Balancing Advisor for subplex EZBT0212
ADNR for subplex EZBT0211 updating zone mvsplex0211.mycorp.com	ADNR for subplex EZBT0212 updating zone mvsplex0212.mycorp.com

Step 2: Configure the name servers that will be updated for the new subplex domains

Configure the name server or name servers with the new zones to represent the new subdomain names created for the new subplexes.

You can use the same name server that was used for the sysplex zones, or another name server or name servers. These name servers must meet the existing requirements for ADNR, as specified in "Step 3: Decide which name server or name servers are to be managed by ADNR" on page 1199. Configure any security requirements in the name server for the new zones, such as access control lists or shared keys. Delegate the new zones as child zones from their parent name servers.

If you are not planning to use the existing sysplex domain name, remove that zone from the name server configuration file on the primary and secondary name servers. Update the parent zone so that it no longer delegates to the zone that you are removing.

Step 3: Define and configure one Advisor per subplex

To configure one Advisor instance per subplex, see "Steps for configuring automated domain name registration" on page 1197.

Configuration includes creating the start procedure JCL, performing the required security product commands (for example, RACF), performing the actions to make the Advisor highly available in case of application or system failure, customizing the TCP/IP profiles, customizing the WLM policies for the Advisor, and so on.

Step 4: Update the Agent configuration files to communicate with the Advisor running in its subplex

Update the security mechanisms that allow the Agent to communicate with its Advisor.

You might need to update the Agent's host_connection statement to point to the Advisor running in its subplex, and the Advisor's agent_id_list statement to allow a connection from this Agent.

If you are using AT-TLS for security, see "Enabling TLS/SSL for ADNR" on page 1197 and "Enabling TLS/SSL for z/OS Load Balancing Advisor (optional)" on page 1164.

Step 5: Define one ADNR application per subplex

If resources in a subplex were previously managed by ADNR, to define one ADNR application per subplex, see the following steps (in "Steps for configuring automated domain name registration" on page 1197):

- "Step 7: Define security server profiles for ADNR" on page 1200
- "Step 8: Configure ADNR to automatically restart in case of application or system failure (optional)" on page 1201
- "Step 9: Configure and start syslogd (optional, but required to have ADNR write log messages and trace data to syslogd)" on page 1202
- "Step 10: Configure one ADNR application per sysplex" on page 1202
- "Step 11: Customize the TCP/IP profiles of the TCP/IP stacks on which ADNR and the LBA applications are to run (optional)" on page 1206

Also, for special considerations when migrating ADNR to a subplexing environment, see:

- "Step 6: Assign the host_group and server_group statements from the sysplex ADNR configuration to the correct subplex domains" on page 1218
- "Step 7: Configure the new ADNR instances to update the name server and zone for its subplex" on page 1219
- "Step 8: Configure the new ADNR instances to communicate with the subplex Advisor" on page 1219

Guideline: In a Common INET (CINET) environment, use stack affinity to ensure that each subplex ADNR instance is using a TCP/IP stack that is in the subplex. For an example of the JCL to use to establish affinity, see adnrproc.sample.

Step 6: Assign the host_group and server_group statements from the sysplex ADNR configuration to the correct subplex domains

For each host_group statement in the original ADNR configuration, identify the subplexes that contain the hosts that are in the group. Create a host_group statement in each new ADNR configuration file of the subplex that contains such a host. Configure the *zone_label* of each host_group statement to point to a zone statement that represents the zone that will represent the subplex. Configure the *dns_label* of the host_group statement to point to a dns statement that represents the name server that owns the specified zone. Create ipaddrlist statements that the host_group statements can reference, such that the IP addresses in them belong to the subplex.

For each server_group statement in the original ADNR configuration, identify the subplexes that contain the server instances that are in the group. Create a server_group statement in each new ADNR configuration file of the subplex that contains such a server instance. Configure the *zone_label* of each server_group statement to point to a zone statement that represents the zone that will represent the subplex. Configure the *dns_label* of the host_group statement to point to a dns statement that represents

the name server that owns the specified zone. Create ipaddrlist statements that the server_group statements can reference, such that the IP addresses in them belong to the subplex.

Step 7: Configure the new ADNR instances to update the name server and zone for its subplex

To configure the dns statements of each subplex ADNR instance so that the correct name server and zone are updated for each subplex, see "Identifying the name servers to update and the zones to be updated in those name servers" on page 1203.

You specified the name server or name servers to be updated in "Step 2: Configure the name servers that will be updated for the new subplex domains" on page 1217. The zones that will be updated for each subplex ADNR were determined in "Step 1: Plan how the new subdomains representing each subplex will fit into your DNS hierarchy" on page 1216.

Step 8: Configure the new ADNR instances to communicate with the subplex Advisor

Configure the gwm statement of each new ADNR instance to communicate with the Advisor that represents that subplex.

Each subplex should have an Advisor that communicates with an ADNR instance, unless there are no resources in that subplex that are to be managed by ADNR. Each new ADNR instance should have a UUID on the uuid statement that is universally unique.

Step 9: Update resolver configuration files (optional)

Users who were using DNS names to connect to resources in the sysplex will be affected by converting to a sysplex subplexing environment. Some DNS names that existed before the migration will no longer exist, and users will need to use new DNS names to connect to the subplex resources.

For example, users that used the name ftp.mvsplex.mycorp.com might now need to use the domain name of the subplex that they are now required to use, such as ftp.mvsplex0212.mycorp.com. You can make some of this transition more transparent to the users by adding the new domain names to the resolver configuration files that are in use throughout your network. Consider adding the new domain names to the resolver configuration SEARCH directive.

Step 10: Start the TCP/IP stacks, Advisor, Agent, ADNR, and target applications that are to be managed by ADNR

See the following steps (in "Steps for configuring automated domain name registration" on page 1197):

- "Step 12: Start the TCP/IP stacks on which ADNR and the LBA applications are to run" on page 1206
- "Step 13: Start the z/OS Load Balancing Advisor and Agent" on page 1207
- "Step 14: Start the target applications that are to be managed by ADNR" on page 1207
- "Step 15: Start the ADNR application" on page 1207

Step 11: Verify that each subplex ADNR is functioning correctly

To verify that each subplex ADNR is functioning as expected, see "Step 16: Verify that the ADNR system is functioning correctly (optional)" on page 1207.

Operating ADNR

When ADNR is operational, you can take the following actions:

- Change the logging level of ADNR to suit your needs (optional)
- Change the ADNR configuration dynamically

- Interpret ADNR display information

Changing the logging level of ADNR

Optionally, you can change the logging level of ADNR to suit your needs. The amount of information that is logged by ADNR can be modified dynamically using the following command:

```
MODIFY procname,DEBUG,LEVEL=debug_level
```

However, modifying the logging level should be carefully considered. The IBM default value of 7 should not be changed unless instructed to do so by an IBM Service representative. For things to consider before modifying this value, see “Customizing optional statements” on page 1206.

Changing the ADNR configuration dynamically

You can change the ADNR configuration dynamically using the ADNR MODIFY *procname*,REFRESH command. For more information, see “Changing the ADNR configuration file” on page 1214.

Interpreting ADNR display information

For information about the MODIFY command for ADNR, including sample ADNR display information and descriptions of the output, see z/OS Communications Server: IP System Administrator's Commands.

Diagnosing problems

ADNR must communicate with a GWM and at least one name server. Problems with this communication are obvious because there is an eventual action message on the console. Problems communicating with a name server are manifested as an eventual action message concerning the name server and separate messages for each zone under the name server, indicating that the zone is unresponsive. Configuration problems with one zone (for example, TSIG key mismatches) are manifested by an eventual action message regarding the name server and a message indicating that the particular zone is unresponsive. In all cases, successful resolution of the underlying problem results in the removal of the eventual action message. For problems communicating with the GWM and for problems communicating with a name server, see z/OS Communications Server: IP Diagnosis Guide.

Regardless of the problem communicating with a name server, ADNR sends health probes in 60-second intervals to test the reachability of the name server and determine whether ADNR and the name server have been compatibly configured. If the problem is transient (for example, a network glitch, or the name server was stopped and quickly restarted), ADNR detects when the name server is again reachable and recovers. If the problem is not transient, it is possible that the ADNR configuration, name server configuration, or both need to be changed. If only the name server configuration needs to be changed, the health probes detect when ADNR and the name server have been compatibly configured and ADNR recovers.

After starting ADNR, after restarting ADNR's GWM, or after issuing a MODIFY REFRESH command, ADNR might be delayed for a period of time before its name servers are updated with the latest sysplex availability information. In all of these cases, after ADNR connects to a GWM, ADNR waits twice the GWM's polling interval before attempting to update its managed name servers. This period of time is called the convergence period. For the definition of convergence period, see “Step 8: Configure ADNR to automatically restart in case of application or system failure (optional)” on page 1201. The purpose of the convergence period is to avoid making premature decisions about which sysplex resources are available or unavailable until all Agents in the sysplex have had sufficient time to receive the data registered by ADNR and report back to the Advisor, and in turn ADNR, on the availability status of those resources. After the convergence period has expired, ADNR updates its managed name servers with the latest sysplex availability information. Henceforth, ADNR updates its managed name servers immediately upon receiving updated status information from the GWM.

Upon starting ADNR or issuing a MODIFY REFRESH command, ADNR resynchronizes itself with its managed name servers. This process consists of performing an asynchronous zone transfer from the name servers of each zone managed by ADNR, and then updating the name servers using this data and

the latest ADNR configuration and sysplex resource availability information. Depending on the number of sysplex resources configured to ADNR and the update_interval value configured on the Advisor, the resynchronization and reconciliation process can take longer than the convergence period when the convergence period is short and the zones are large. When the zone status reaches SYNCHRONIZED, this process has completed and the zone should reflect the latest status information from the sysplex.

ADNR configuration example

This topic includes a specific configuration example of ADNR. This example assumes that the z/OS Load Balancing Advisor solution is installed.

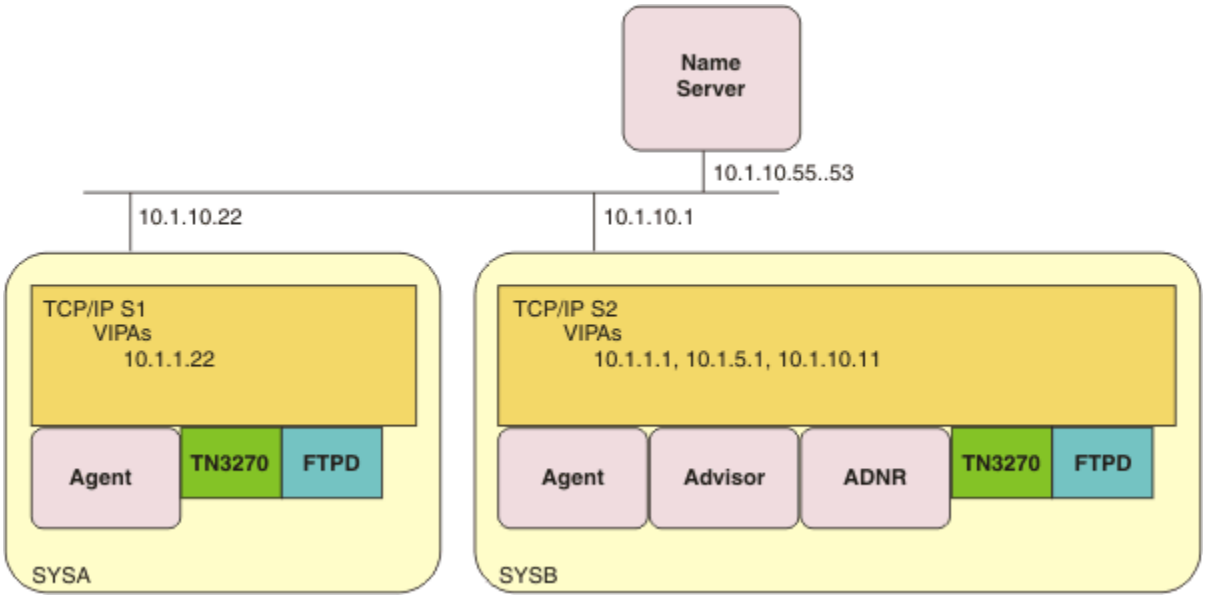


Figure 150. ADNR configuration example

In Figure 150 on page 1221, two systems, SYSA and SYSB, are in a sysplex. The z/OS Load Balancing Advisor solution is configured and deployed in the sysplex, with the Advisor and ADNR running on SYSB and an Agent running on both SYSA and SYSB. The TN3270E Telnet server and FTPD are active on both systems. ADNR manages a name server in the network.

The following example shows an ADNR configuration file. The numbers in the left margin were added for annotation purposes and are used in the description that follows the example. This example configuration does not use subplexing or AT-TLS.

```
001 debug_level             7               # Error, Warning, Event
002
003 uuid                    mycorp_sysplex_adnr
004
005 dns                     network_name_server # Label used by other stmts
006                                         #  and commands
007 {
008   dns_id                10.1.10.55      # Network name server
009
010   zone                  mvsplex.mycorp.com_zone
011                                         # Label used by other stmts
012                                         #  and commands
013   {
014     domain_suffix       mvsplex.mycorp.com
015                                         # Zone name in name server
016   } # end of zone
017
018 } # end of dns
019
020 gwm                     z/os_lba_advisor
021 {
022   gwm_id                10.1.5.1..3860  # LBA lb_connection_v4 address
023   host_connection_addr  10.1.10.11      # Local address
024 } # end of gwm
025
```

```
026 host_group              production_sysplex
027                                     # Addrs available to intranet
028 {
029   host_group_name       prodplex        # Prepended to domain suffix
030   dns                   network_name_server # Name server to update
031   zone                  mvsplex.mycorp.com_zone
032                                     # Determines zone suffix for
033                                     #  this group
034                                     #  (mvsplex.mycorp.com)
035
036   member                            # sysa.mvsplex.mycorp.com,
037                                     #  and
038                                     #  prodplex.mvsplex.mycorp.com
039   {
040     host_name           sysa            # Prepended to domain suffix
041     ipaddrlist          sysa_vipa_addrs
042     ipaddrlist          sysa_non_vipa_addrs
043   }
044
045   member                            # sysb.mvsplex.mycorp.com,
046                                     #  and
047                                     #  prodplex.mvsplex.mycorp.com
048   {
049     host_name           sysb            # Prepended to domain suffix
050     ipaddrlist          sysb_vipa_addrs
051     ipaddrlist          sysb_non_vipa_addrs
052   }
053
054 } # end of host_group
055
056 ipaddrlist              sysa_vipa_addrs
057 {
058   ipaddr                10.1.1.22
059 } # end of ipaddrlist
060
061 ipaddrlist              sysa_non_vipa_addrs
062 {
063   ipaddr                10.1.10.22      # OSA on sysa
064 } # end of ipaddrlist
065
066 ipaddrlist              sysb_vipa_addrs
067 {
068   ipaddr                10.1.1.1
069 } # end of ipaddrlist
070
071 ipaddrlist              sysb_non_vipa_addrs
072 {
073   ipaddr                10.1.10.1       # OSA on sysb
074 } # end of ipaddrlist
075
076 server_group            tn3270_group    # TN3270E Telnet servers
077 {
078
079   port                  23              # TN3270E Telnet server port
080   protocol              TCP             # Protocol for this port
081   server_group_name     ztelnet         # Prepended to domain suffix
082   dns                   network_name_server # Name server to update
083   zone                  mvsplex.mycorp.com_zone
084                                     # Determines zone suffix for
085                                     #  this group
086                                     #  (mvsplex.mycorp.com)
087   member                # telnetprimary.ztelnet.mvsplex.mycorp.com
088                         #  and,
089                         #  ztelnet.mvsplex.mycorp.com
090   {
091     server_name         telnetprimary   # Prepended to
092                                     #  server_group_name.domain
093                                     #  suffix
094     ipaddrlist          sysa_vipa_addrs
095   }
096
097   member                # telnetsecondary.ztelnet.mvsplex.mycorp.com
098                         #  and,
099                         #  ztelnet.mvsplex.mycorp.com
100   {
101     server_name         telnetsecondary # Prepended to
102                                     #  server_group_name.domain
103                                     #  suffix
104     ipaddrlist          sysb_vipa_addrs
105   }
106 } # end of server_group
107
```

```
108 server_group            ftp_group       # FTP daemons
109 {
110
111   port                  21              # FTP port
112   protocol              TCP             # Protocol for this port
113   server_group_name     zftp            # Prepended to domain suffix
114   dns                   network_name_server
115                                         # Name server to update
116   zone                  mvsplex.mycorp.com_zone
117                                         # Determines zone suffix for
118                                         #  this group
119                                         #  (mvsplex.mycorp.com)
120   member                # zftp.mvsplex.mycorp.com
121   {
122     ipaddrlist          sysa_vipa_addrs
123     ipaddrlist          sysb_vipa_addrs
124   }
125 } # end of server_group
```

- Line 1:

  In this example ADNR configuration file, the debug level is set to 7 in the optional debug_level statement on line 1. The value of 7 is the default value, so this statement is redundant but is shown for completeness. At the default level of 7, messages are written to the log if they are at error, warning, or event level. Messages at other debug levels, such as info or debug, are not written to the log file.

- Line 3:

  The uuid statement on line 3 uniquely identifies this ADNR from all other ADNR instances and external load balancers. In this example, it is a character string giving some useful information about the location of the ADNR application.

- Line 5:

  There is one name server that ADNR will manage in this example, which is represented by the dns statement on line 5. The statement is given a label of network_name_server that is used for references from other statements, like host_group and server_group, and might be used in display commands. This dns statement is referenced on lines 30, 82, and 114.

- Line 8:

  The dns_id keyword within the dns statement on line 8 tells ADNR how to reach the name server. The name server configured is at address 10.1.10.55, listening on port 53. The port is not explicitly specified because the default port is being used.

- Line 10:

  The zone keyword within the dns statement on line 10 indicates which zones ADNR manages at the name server indicated on the dns_id keyword. There can be multiple zone keywords in this statement. However, in this example, only one zone is managed by ADNR in this name server. The zone keyword has the label mvsplex.mycorp.com_zone, which is used for references from other statements like host_group and server_group. In this example, the zone is referenced on lines 31, 83, and 116.

- Line 14:

  The domain_suffix keyword within the zone keyword on line 14 assigns the domain suffix to be used for all updates to the zone. In this example, the domain suffix of mvsplex.mycorp.com is appended to the host names that ADNR creates when it adds resource records to the name server. In this example, one of the resource records added to the name server is ztelnet.mvsplex.mycorp.com. Because the zone keyword containing this domain suffix is under the dns statement representing the name server at 10.1.10.55, this implies that the name server at 10.1.10.55 is configured as the primary master name server for the mvsplex.mycorp.com domain.

- Line 20:

  The gwm statement on line 20 describes how ADNR communicates with the GWM. The gwm statement has a label of z/os_lba_advisor.

- Line 22:

The gwm_id keyword on line 22 identifies the IP address and port on which the GWM is listening for connections from load balancers. In this example, the GWM is located at 10.1.5.1 and is listening on port 3860 for load balancer connections. The port of 3860 on the gwm_id keyword does not need to be specified because that is the default port for this keyword. However, it is explicitly coded in this example for clarity. ADNR establishes a long-lived TCP connection to the GWM.

- Line 23:

  The host_connection_addr keyword of the gwm statement on line 23 identifies the source IP address that ADNR uses when connecting to the GWM. The host_connection_addr keyword is optional if you are using AT-TLS. In this example, 10.1.10.11 is used. The GWM might require configuration to enable a load balancer to connect from this address. Because the Advisor is acting as the GWM, it requires either that the source IP address of all load balancers that connect to it appear in an access control list or that AT-TLS is used. This list is defined by the lb_id_list statement in the Advisor. Therefore, in this example, 10.1.10.11 must appear in the Advisor's lb_id_list statement to enable ADNR to connect to it.

- Line 26:

  The host_group statement defines a set of IP addresses to potentially add to the ADNR-managed name server, which represents the hosts in the sysplex. They are used to map names for the hosts to IP addresses in the name server. This is one of the traditional uses of resource records in a name server. In this example, production_sysplex is the label for this host_group on line 26. This label can be used in display commands. There can be multiple host_group statements in an ADNR configuration file, although only one is shown in this example.

- Lines 29 and 31:

  The host_group_name keyword identifies the DNS host name that is associated with all IP addresses identified in this host_group statement. The value of the host_group_name keyword, `prodplex` on line 29 in this example, is prepended to the domain suffix identified by the zone keyword of this host_group statement on line 31. In this example, all IP addresses identified by the ipaddrlist statements in this host_group statement have the potential to be added to the name server with the name prodplex.mvsplex.mycorp.com. The IP addresses that are actually added to the name server with this name are the intersection of all of the IP addresses identified in all ipaddrlist keywords in this host_group statement, with the addresses that exist in the union of all home lists in the sysplex. This assumes that an Agent is running on all systems in the sysplex. If an address is removed from the home list of a system in the sysplex (for example, using VARY TCPIP,,OBEYFILE), all resource records associated with that IP address are update-deleted from the ADNR-managed name server. Similarly, if an IP address is added to the home list of a system in the sysplex and that address is already represented in this host_group statement, one or more resource records with that IP address are update-added to the ADNR-managed name server. The number of resource records added depends upon ADNR configuration. Assuming that all IP addresses configured to this host group actually exist in the sysplex, the following resource records are added to the ADNR-managed name server (TTLs are omitted):

  ```
  prodplex.mvsplex.mycorp.com   IN A  10.1.1.22
  prodplex.mvsplex.mycorp.com   IN A  10.1.10.22
  prodplex.mvsplex.mycorp.com   IN A  10.1.1.1
  prodplex.mvsplex.mycorp.com   IN A  10.1.10.1
  ```

  Thus, the DNS name prodplex.mvsplex.mycorp.com can be used by a client to reach any system in the sysplex. Because there is no guarantee that a server is listening on each of these addresses, this DNS name should not be used by clients to connect to servers in the sysplex. This DNS name is probably most useful for utilities like ping and traceroute.

- Line 30:

  The dns keyword in the host_group statement on line 30 identifies the name server that is to be updated for this host group. The dns keyword identifies this name server by the label network_name_server, which is a label on a dns statement (line 5 in this example). Thus, in this example, the names for this host_group statement are added to the name server located at 10.1.10.55.

- Line 31:

The zone keyword of the host_group statement on line 31 identifies the zone to which the names for this host group are to be added. In this example, the label mvsplex.mycorp.com_zone matches a zone keyword label in the dns statement on line 10. Because multiple zone keywords are possible in a dns statement, the zone within the dns statement must be identified from all host_group and server_group statements. The zone keyword in the host_group statement is how the domain suffix is determined for this host_group statement.

- Lines 36 and 45:

  This host_group statement contains two member definitions on lines 36 and 45. Both of these member definitions are named members because they contain a host_name keyword. An unnamed member has no host_name keyword. You can code one unnamed member per host_group. However, if one is not coded, a virtual unnamed member is created for you. A virtual unnamed member contains the union of all IP addresses coded in named members within that host group. An explicitly coded unnamed member also contains the union of all IP addresses coded in the named members that belong to that host group, in addition to any IP addresses coded in the unnamed member itself. An explicitly coded unnamed member might be useful for configuring movable dynamic VIPAs that do not always belong to a specific host.

- Line 40:

  The first member definition in this host_group statement contains the host_name keyword with a value of sysa. All IP addresses that are coded in this member belong to the sysa system. The value sysa of the host_name keyword on line 40 is prepended to the domain suffix for this host group to create the name sysa.mvsplex.mycorp.com. Assuming that all IP addresses configured to this member actually exist on sysa, the following resource records are added to the ADNR-managed name server (TTLs are omitted):

  ```
  sysa.mvsplex.mycorp.com    IN A  10.1.1.22
  sysa.mvsplex.mycorp.com    IN A  10.1.10.22
  ```

  Thus, the DNS name sysa.mvsplex.mycorp.com can be used to specifically connect to the sysa system. This name might be useful for ping or traceroute.

- Line 45:

  The second member definition in the host_group statement on line 45 represents IP addresses that are to be associated with the sysb system, like the first member definition associated IP addresses with the sysa system. Assuming that all IP addresses configured to this member actually exist on sysb, the following resource records are added to the ADNR-managed name server (TTLs are omitted):

  ```
  sysb.mvsplex.mycorp.com    IN A  10.1.1.1
  sysb.mvsplex.mycorp.com    IN A  10.1.10.1
  ```

  Similar to the first member definition, the sysb.mvsplex.mycorp.com DNS name can be used to specifically connect to the sysb system.

- Lines 41, 42, 50, and 51:

  The two ipaddrlist keywords in each of the two members of the host_group statement (lines 41, 42, 50, and 51) specify the IP addresses that belong to this host group. As this example shows, multiple ipaddrlist keywords are allowed and the group represents the union of all ipaddrlist statements in the host_group statement. Each ipaddrlist statement can contain multiple IP addresses, although this example has only one IP address per ipaddrlist statement. In this host_group statement, the two ipaddrlist keywords in the first member (lines 41 and 42) reference ipaddrlist statements with the labels sysa_vipa_addrs and sysa_non_vipa_addrs at lines 56 and 61, and the two ipaddrlist keywords in the second member (lines 50 and 51) reference the ipaddrlist statements with the labels sysb_vipa_addrs and sysb_non_vipa_addrs at lines 66 and 71. Thus, the IP addresses 10.1.1.22 and 10.1.10.22 are associated with this host group, as are the IP addresses 10.1.1.1 and 10.1.10.1, as reflected in the IP addresses associated with the name prodplex.mvsplex.mycorp.com as previously shown. Furthermore, the IP addresses of the first member, 10.1.1.22 and 10.1.10.22, are associated with system sysa, as reflected in the IP addresses associated with the name sysa.mvsplex.mycorp.com, and the IP

addresses in the second member, 10.1.1.1 and 10.1.10.1, are associated with system sysb, as reflected in the IP addresses associated with the name sysb.mvsplex.mycorp.com.

- Lines 56, 61, 66, and 71:

  This example shows four ipaddrlist statements, each with one IP address, at lines 56, 61, 66, and 71. There can be multiple ipaddrlist statements and each can have multiple IP addresses. Each ipaddrlist statement has a label that is referenced from host_group and server_group statements. The labels in this example are sysa_vipa_addrs, sysa_non_vipa_addrs, sysb_vipa_addrs, and sysb_non_vipa_addrs. This example shows only IPv4 addresses, but IPv6 addresses can also be coded in ipaddrlist statements. The sample ADNR configuration file shows examples of IPv6 addresses coded in ipaddrlist statements. The sample ADNR configuration file can be found in the EZBADNRC member of SEZAINST.

- Lines 58, 63, 68, and 73:

  The ipaddr keyword on lines 58, 63, 68, and 73 identifies a single IP address to register. This example shows only IPv4 addresses, but IPv6 addresses can also be coded in ipaddrlist statements. The sample ADNR configuration file shows examples of IPv6 addresses coded in ipaddrlist statements. The sample ADNR configuration file can be found in the EZBADNRC member of SEZAINST.

- Lines 76 and 108:

  This example shows two server_group statements; one at line 76 to represent the TN3270E Telnet servers in the sysplex, and the other at line 108 to represent the FTP daemons in the sysplex. The server_group statements are used to create DNS names in the name server that map to IP addresses, which can actually be used to reach an active instance of the server.

- Line 76:

  The first server_group statement on line 76 is given the label tn3270_group. This label can be used in display commands. This group represents TN3270E Telnet servers in the sysplex.

- Line 79:

  The port keyword on line 79 in the first server_group statement indicates the port that the TN3270E Telnet servers listen on, which is 23 in this example.

- Line 80:

  The protocol keyword on line 80 in the first server_group statement indicates the protocol of the listening application (TCP or UDP). Because the TN3270E Telnet server is a TCP application, the value for this keyword is TCP.

- Line 81:

  The server_group_name keyword on line 81 identifies the DNS host name that is associated with all IP addresses identified in this server_group statement. The value of the server_group_name keyword, `ztelnet` in this example, is prepended to the domain suffix identified by the zone keyword of this server_group statement. In this example, all IP addresses identified by the ipaddrlist statements in this server_group statement have the potential to be added to the name server with the name ztelnet.mvsplex.mycorp.com. The IP addresses that are actually added to the name server with this name are the intersection of all of the IP addresses identified in all ipaddrlist keywords in this server_group statement, with the addresses on which the server instances in the sysplex are available. In the case of TCP applications, this is the set of IP addresses that the servers are listening on, and for UDP applications, this is the set of IP addresses to which the servers are bound. This assumes that an Agent is running on all systems in the sysplex. The TN3270E Telnet server listens on INADDR_ANY, so in effect, it listens on all IP addresses available. If a server instance is stopped (or abnormally terminates), all resource records associated with that server instance are update-deleted from the ADNR-managed name server. Similarly, if a new server instance in that group is started and an IP address on which that server instance is available is already coded in the server_group statement, one or more resource records with that IP address are update-added to the ADNR-managed name server. The number of resource records added depends upon ADNR configuration. Assuming the TN3270E

Telnet servers are running on sysa and sysb, the following resource records are added to the ADNR-managed name server (TTLs are omitted):

```
ztelnet.mvsplex.mycorp.com    IN A  10.1.1.22
ztelnet.mvsplex.mycorp.com    IN A  10.1.1.1
```

Thus, the DNS name ztelnet.mvsplex.mycorp.com can be used to reach any available TN3270E Telnet server in the sysplex.

- Lines 82 and 83:

  On lines 82 and 83, the dns and zone keywords of the server_group statement serve the same purpose as those keywords on the host_group statement.

- Lines 87 and 97:

  There are two named members defined in the server_group statement on lines 87 and 97. Named members contain a server_name keyword, but unnamed members do not. As in host_group statements, you can code one unnamed member per server group. However, if one is not coded, a virtual unnamed member is created for you. A virtual unnamed member contains the union of all IP addresses coded in named members within that server group. An explicitly coded unnamed member also contains the union of all IP addresses coded in the named members that belong to that server group, in addition to any IP addresses coded in the unnamed member itself. The IP addresses in the unnamed member, whether explicitly coded or not, are associated with the group name. In this example, the IP addresses of the unnamed member for this server_group statement are associated with the name ztelnet.mvsplex.mycorp.com. Coding an explicit unnamed member without any named members might suffice for servers where affinity to a particular instance is not needed. Also, if an application instance is freely movable from one system to another, or from one TCP/IP stack to another, coding an unnamed member without any named members might be acceptable. In this case, code the DVIPAs on which the application could be reached in the unnamed member. If, however, it is important for clients to be able to connect to specific instances of a server, you must code separate named members for each instance.

- Line 91:

  The first member definition in this server_group statement contains the server_name keyword with a value of `telnetprimary` at line 91. All IP addresses that are coded in this member are IP addresses that one instance of the TN3270E Telnet server could potentially be listening on. The value `telnetprimary` of the server_name keyword is prepended to the DNS name created for the server group to create the name telnetprimary.ztelnet.mvsplex.mycorp.com. Assuming that the IP address configured to this member actually exists on sysa, and that the TN3270E Telnet server is active on this system, the following resource record is added to the ADNR-managed name server (TTLs are omitted):

  ```
  telnetprimary.ztelnet.mvsplex.mycorp.com    IN A  10.1.1.22
  ```

  Thus, the DNS name telnetprimary.ztelnet.mvsplex.mycorp.com can be used to specifically connect to the TN3270E Telnet server instance on sysa.

- Line 97:

  The second member definition in this server group on line 97 is similar to the first, with the exception that this member identifies an application instance on system sysb instead of the one on system sysa. Assuming that the IP address configured to this member actually exists on sysb, and that the TN3270E Telnet server is active on this system, the following resource record is added to the ADNR-managed name server (TTLs are omitted):

  ```
  telnetsecondary.ztelnet.mvsplex.mycorp.com  IN A  10.1.1.1
  ```

  Thus, the DNS name telnetsecondary.ztelnet.mvsplex.mycorp.com can be used to specifically connect to the TN3270E Telnet server instance on sysb.

- Line 108:

  The second server_group statement on line 108 is similar to the server_group statement for the TN3270E Telnet servers, except that it represents FTP daemons in the sysplex. Thus, the value of the port keyword specifies the port number on which the FTP daemon listens. The server_group_name

value specifies a name that is appropriate to a group of equivalent FTP daemons. This server group has one unnamed member and no named members. Thus, the resource records that can potentially be added to the name server are as follows:

```
zftp.mvsplex.mycorp.com    IN A  10.1.1.22
zftp.mvsplex.mycorp.com    IN A  10.1.1.1
```

ADNR display examples

The displays in this topic use the example configuration described in “ADNR configuration example” on page 1221. The numbers in the left margin were added for annotation purposes and are used in the descriptions of the displays. Only a subset of the possible types of displays are shown in this example. For more examples of the MODIFY command for ADNR, see z/OS Communications Server: IP System Administrator's Commands.

The following display shows information about the gwm:

```
01 F ADNR,DISP,GWM,DETAIL
02 EZD1254I GWM DETAIL
03 GWM LABEL        : Z/OS_LBA_ADVISOR
04  GWM STATUS      : CONVERGENCE_PENDING
05  GWM TIMESTAMP   : 10/19/05 18:32:52
06  GWM IPADDR..PORT: 10.1.5.1..3860
07  LOCAL IPADDR    : 10.1.10.11
08  UUID            : MYCORP_SYSPLEX_ADNR
09  UPDATE INTERVAL : 60
10  LAST UPDATE     : N/A
11 1 OF 1 RECORDS DISPLAYED
```

The following comments refer to the previous display:

- Line 3:

  The GWM LABEL field on line 3 is taken from the gwm statement label on line 20 of the example ADNR configuration file. All labels and names in ADNR displays are displayed in uppercase regardless of how they appear in the ADNR configuration file.

- Line 4:

  The GWM STATUS on line 4 is CONVERGENCE_PENDING, which is a transient state meaning that the ADNR host_group and server_group information has successfully been registered with the GWM, and ADNR is waiting for a timer to expire before updating the name servers with information from this GWM.

- Line 5:

  The GWM TIMESTAMP field on line 5 shows the date and time when the most recent update of the GWM STATUS field occurred.

- Line 6:

  The GWM IPADDR..PORT field shows the IP address and port of the GWM. These values are taken from the gwm_id keyword of the gwm statement on line 22 of the example configuration file.

- Line 7:

  The LOCAL IPADDR field is the source IP address that ADNR uses when communicating with the GWM. The value of 10.1.10.11 was taken from the host_connection_addr keyword on line 23 of the example configuration file

  .

- Line 8:

  The UUID field value on line 8 is taken from the uuid statement on line 3 of the example configuration file.

- Line 9:

  The update interval of 60 seconds is determined by GWM configuration. For the z/OS Load Balancing Advisor, this value is determined by the Advisor's update_interval statement.

- Line 10:

  The LAST UPDATE field on line 10 shows the most recent date and time that status information was received from the GWM on any of the sysplex resources that ADNR registered with the GWM. The value N/A indicates that no status updates have been received yet.

- Line 11:

  The number of records displayed on line 11 indicates the number of GWMs actually displayed. Currently, this value is always 1, because only one GWM is supported.

The following display was created after the convergence timer expired. The GWM's status changes to GWM_ACTIVE on line 4. Any status updates from the GWM will now be reflected in the name servers, provided that the name servers are reachable and configured correctly.

```
01 F ADNR,DISP,GWM,DETAIL
02 EZD1254I GWM DETAIL
03 GWM LABEL        : Z/OS_LBA_ADVISOR
04  GWM STATUS      : GWM_ACTIVE
05  GWM TIMESTAMP   : 10/19/05 18:34:55
06  GWM IPADDR..PORT: 10.1.5.1..3860
07  LOCAL IPADDR    : 10.1.10.11
08  UUID            : MYCORP_SYSPLEX_ADNR
09  UPDATE INTERVAL : 60
10  LAST UPDATE     : 10/19/05 18:35:29
11 1 OF 1 RECORDS DISPLAYED
```

The following display summary shows the host groups and server groups that were defined:

```
01 F ADNR,DISP,GWM,GROUPS
02 EZD1254I GWM GROUP SUMMARY
03 GWM LABEL        : Z/OS_LBA_ADVISOR
04  GWM STATUS      : GWM_ACTIVE
05  GROUP LABEL     : PRODUCTION_SYSPLEX
06   GROUP NAME     : PRODPLEX
07  GROUP LABEL     : TN3270_GROUP
08   GROUP NAME     : ZTELNET
09  GROUP LABEL     : FTP_GROUP
10   GROUP NAME     : ZFTP
11 3 OF 3 RECORDS DISPLAYED
```

The following comments refer to the previous display:

- Lines 3 and 4:

  Lines 3 and 4 show the same information as the DETAIL display.

- Lines 5, 7, and 9:

  The GROUP LABEL fields on lines 5, 7, and 9 show the labels of the groups that were defined to ADNR on lines 26, 76, and 108, respectively, in the example ADNR configuration file.

- Lines 6, 8, and 10:

  The GROUP NAME fields on lines 6, 8, and 10 show the names from the host_group_name or server_group_name keywords on lines 29, 81, and 113, respectively, of the example ADNR configuration file.

- Line 11:

  The number of records displayed on line 11 indicates the number of groups actually displayed. When the number of groups defined exceeds the MAX parameter value (or the default of 100) of the MODIFY ADNR,DISPLAY command, the difference between the numbers on this line indicates the number of groups that were omitted from the display as a result of the MAX parameter value.

The following display shows the details of one group, and demonstrates how a label can be used in a display command:

```
01 F ADNR,DISP,GWM,GROUPS,GROUPID=PRODUCTION_SYSPLEX,DETAIL
02 EZD1254I GWM GROUP DETAIL
03 GWM LABEL        : Z/OS_LBA_ADVISOR
04  GWM STATUS      : GWM_ACTIVE
05  GWM TIMESTAMP   : 10/20/05 13:15:19
```

```
06  GWM IPADDR..PORT: 10.1.5.1..3860
07  LOCAL IPADDR    : 10.1.10.11
08  UUID            : MYCORP_SYSPLEX_ADNR
09  UPDATE INTERVAL : 60
10  LAST UPDATE     : 10/20/2005 13:17:21
11  GROUP LABEL     : PRODUCTION_SYSPLEX
12   GROUP NAME     : PRODPLEX
13   GROUP TYPE     : HOST
14   DNS LABEL      : NETWORK_NAME_SERVER
15   ZONE LABEL     : MVSPLEX.MYCORP.COM_ZONE
16   MEMBER HOSTNAME:
17    IPADDR        : 10.1.1.22
18     AVAIL        : YES
19     FLAGS        :
20     UPDATE COUNT : 1
21    IPADDR        : 10.1.10.22
22     AVAIL        : YES
23     FLAGS        :
24     UPDATE COUNT : 1
25    IPADDR        : 10.1.1.1
26     AVAIL        : YES
27     FLAGS        :
28     UPDATE COUNT : 1
29    IPADDR        : 10.1.10.1
30     AVAIL        : YES
31     FLAGS        :
32     UPDATE COUNT : 1
33   MEMBER HOSTNAME: SYSA
34    IPADDR        : 10.1.1.22
35     AVAIL        : YES
36     FLAGS        :
37     UPDATE COUNT : 1
38    IPADDR        : 10.1.10.22
39     AVAIL        : YES
40     FLAGS        :
41     UPDATE COUNT : 1
42   MEMBER HOSTNAME: SYSB
43    IPADDR        : 10.1.1.1
44     AVAIL        : YES
45     FLAGS        :
46     UPDATE COUNT : 1
47    IPADDR        : 10.1.10.1
48     AVAIL        : YES
49     FLAGS        :
50     UPDATE COUNT : 1
51 1 OF 1 RECORDS DISPLAYED
```

The following comments refer to the previous display:

- Lines 3–10:

  Lines 3 through 10 show information described in previous display examples.

- Line 11:

  The GROUP LABEL field on line 11 matches the value on the GROUPID keyword on the MODIFY DISPLAY command.

- Line 13:

  The GROUP TYPE field on line 13 indicates that this group was defined as a host_group versus a server_group.

- Line 14:

  The DNS LABEL field on line 14 is the label referenced by the dns keyword on line 30 of the example ADNR configuration file, which indicates which name server is to be updated with the information from this group.

- Line 15:

  The ZONE LABEL field on line 15 is the label referenced by the zone keyword on line 31 of the example ADNR configuration file, which indicates which zone in the name server is to be updated with the information from this group.

- Line 16:

The MEMBER HOSTNAME field on line 16 is blank, indicating that this is the unnamed member for this group. Because an unnamed member was not explicitly configured for this group, this member represents the virtual unnamed member that is created for you.

- Lines 17, 21, 25, and 29:

  The IPADDR values on lines 17, 21, 25, and 29 represent the IP addresses that are automatically added to the unnamed member. These four addresses represent the union of all IP addresses coded in all members of this group.

- Lines 18, 22, 26, and 30:

  The AVAIL flags on lines 18, 22, 26, and 30 all indicate that the GWM has reported these resources as available. For IPADDRs in a host group such as this, this means the Agent owning these IP addresses is active and these addresses exist in a home list in the sysplex.

- Lines 19, 23, 27, and 31:

  The FLAGS field for each of the IP addresses on lines 19, 23, 27, and 31 is empty. If the AVAIL flags were NO for any of these IP addresses, the FLAGS field would give an indication as to why. For a list of the possible flags that can appear for the MODIFY command for ADNR, and their explanations, see z/OS Communications Server: IP System Administrator's Commands.

- Lines 20, 24, 28, and 32:

  The UPDATE COUNT fields on lines 20, 24, 28, and 32 indicate the number of times that the availability status, as sent from the GWM, has changed for these IP addresses.

- Lines 33 and 42:

  The MEMBER HOSTNAME fields on lines 33 and 42 represent the named members that were configured in the group at lines 36 and 45 of the example ADNR configuration file. Only the IP addresses explicitly coded to these members appear in the display of these members, in contrast to the unnamed member. The remainder of the fields in these named members are equivalent to the information displayed in the unnamed member.

The following display shows information about the name server that ADNR is updating:

```
01 F ADNR,DISP,DNS,DETAIL
02 EZD1254I DNS DETAIL
03 DNS LABEL        : NETWORK_NAME_SERVER
04  DNS STATUS      : ACTIVE
05  DNS IPADDR..PORT: 10.1.10.55..53
06  ZONES DEFINED   : 1
07  ZONES ACTIVE    : 0
08 1 OF 1 RECORDS DISPLAYED
```

The following comments refer to the previous display:

- Line 3:

  The DNS LABEL field value NETWORK_NAME_SERVER is taken from the dns statement on line 5 of the example configuration file. For an explanation of all of the DNS states displayed on the MODIFY command for ADNR, see z/OS Communications Server: IP System Administrator's Commands.

- Line 5:

  The DNS IPADDR..PORT field value of 10.1.10.55..53 indicates that ADNR communicates with a name server at address 10.1.10.55 on port 53. This IP address and port are taken from the dns_id keyword on line 8 of the example configuration file.

- Line 6:

  The ZONES DEFINED field on line 6 indicates that there is one zone configured under this dns statement in the ADNR configuration file.

- Line 7:

  The ZONES ACTIVE field on line 7 indicates that the zone defined under this dns statement cannot yet be updated by ADNR. In a steady state environment, the number of active zones should match the number of defined zones. If the numbers do not match, this could be an indication of a configuration

problem, a network problem, or the zone could be in the middle of the resynchronization process. For more information on the resynchronization process, see “Operating ADNR” on page 1219.

- Line 8:

  The number of records displayed on line 8 indicates the number of DNSs actually displayed. When the number of DNSs defined exceeds the MAX parameter value (or the default of 100) of the MODIFY ADNR,DISPLAY command, the difference between the numbers on this line indicates the number of DNSs that were omitted from the display as a result of the MAX parameter value.

The following display shows summary information about the zones that ADNR is updating:

```
01 F ADNR,DISP,DNS,ZONES,SUMMARY
02 EZD1254I DNS ZONE SUMMARY
03 DNS LABEL         : NETWORK_NAME_SERVER
04  DNS STATUS       : ACTIVE
05  ZONE LABEL       : MVSPLEX.MYCORP.COM_ZONE
06   ZONE STATUS     : SYNCHRONIZED
07 1 OF 1 RECORDS DISPLAYED
```

The following comments refer to the previous display:

- Line 3:

  The DNS LABEL field on line 3 indicates that information for a specific name server and its zones immediately follows. The field value is taken from the label on the dns statement on line 5 of the example configuration file. For an explanation of all of the DNS states displayed on the MODIFY command for ADNR, see z/OS Communications Server: IP System Administrator's Commands.

- Line 5:

  The ZONE LABEL field on line 5 indicates that information for the indicated zone follows. The value for this field is taken from the label on the zone keyword (line 10 of the example configuration file) of the dns statement.

- Line 6:

  The ZONE STATUS field on line 6 shows the state of the zone. The value of SYNCHRONIZED is the expected, steady-state value for a zone. For an explanation of all of the zone states displayed on the MODIFY command for ADNR, see z/OS Communications Server: IP System Administrator's Commands.

- Line 7:

  The number of records displayed on line 7 indicates the number of zones actually displayed. When the number of zones defined exceeds the MAX parameter value (or the default of 100) of the MODIFY ADNR,DISPLAY command, the difference between the numbers on this line indicates the number of zones that were omitted from the display as a result of the MAX parameter value.

The following display shows the status of the resource records that ADNR manages in the name server:

```
001 F ADNR,DISP,DNS,ZONES,DETAIL
002 EZD1254I DNS ZONE DETAIL
003 DNS LABEL         : NETWORK_NAME_SERVER
004  DNS STATUS       : ACTIVE
005  DNS IPADDR..PORT: 10.1.10.55..53
006  ZONES DEFINED    : 1
007  ZONES ACTIVE     : 1
008  ZONE LABEL       : MVSPLEX.MYCORP.COM_ZONE
009   ZONE STATUS     : SYNCHRONIZED
010   DOMAIN SUFFIX   : MVSPLEX.MYCORP.COM.
011   ZONE TIMESTAMP  : 10/20/05 14:42:05
012   TSIG FLAGS      :
013   DNS RR LABEL    : PRODPLEX
014    DNS RR STATUS  : PRESENT
015    TTL            : 60
016    CLASS          : IN
017    TYPE           : A
018    RDATA          : 10.1.1.22
019    GWM LABEL      : Z/OS_LBA_ADVISOR
020    GROUP LABEL    : PRODUCTION_SYSPLEX
021    LAST UPDATE    : 10/20/05 14:42:02
022   DNS RR LABEL    : PRODPLEX
023    DNS RR STATUS  : PRESENT
```

```
024    TTL           : 60
025    CLASS         : IN
026    TYPE          : A
027    RDATA         : 10.1.1.1
028    GWM LABEL     : Z/OS_LBA_ADVISOR
029    GROUP LABEL   : PRODUCTION_SYSPLEX
030    LAST UPDATE   : 10/20/05 14:42:02
031   DNS RR LABEL   : PRODPLEX
032    DNS RR STATUS : PRESENT
033    TTL           : 60
034    CLASS         : IN
035    TYPE          : A
036    RDATA         : 10.1.10.1
037    GWM LABEL     : Z/OS_LBA_ADVISOR
038    GROUP LABEL   : PRODUCTION_SYSPLEX
039    LAST UPDATE   : 10/20/05 14:42:02
040   DNS RR LABEL   : PRODPLEX
041    DNS RR STATUS : PRESENT
042    TTL           : 60
043    CLASS         : IN
044    TYPE          : A
045    RDATA         : 10.1.10.22
046    GWM LABEL     : Z/OS_LBA_ADVISOR
047    GROUP LABEL   : PRODUCTION_SYSPLEX
048    LAST UPDATE   : 10/20/05 14:42:02
049   DNS RR LABEL   : SYSA
050    DNS RR STATUS : PRESENT
051    TTL           : 60
052    CLASS         : IN
053    TYPE          : A
054    RDATA         : 10.1.1.22
055    GWM LABEL     : Z/OS_LBA_ADVISOR
056    GROUP LABEL   : PRODUCTION_SYSPLEX
057    LAST UPDATE   : 10/20/05 14:42:02
058   DNS RR LABEL   : SYSA
059    DNS RR STATUS : PRESENT
060    TTL           : 60
061    CLASS         : IN
062    TYPE          : A
063    RDATA         : 10.1.10.22
064    GWM LABEL     : Z/OS_LBA_ADVISOR
065    GROUP LABEL   : PRODUCTION_SYSPLEX
066    LAST UPDATE   : 10/20/05 14:42:02
067   DNS RR LABEL   : SYSB
068    DNS RR STATUS : PRESENT
069    TTL           : 60
070    CLASS         : IN
071    TYPE          : A
072    RDATA         : 10.1.1.1
073    GWM LABEL     : Z/OS_LBA_ADVISOR
074    GROUP LABEL   : PRODUCTION_SYSPLEX
075    LAST UPDATE   : 10/20/05 14:42:02
076   DNS RR LABEL   : SYSB
077    DNS RR STATUS : PRESENT
078    TTL           : 60
079    CLASS         : IN
080    TYPE          : A
081    RDATA         : 10.1.10.1
082    GWM LABEL     : Z/OS_LBA_ADVISOR
083    GROUP LABEL   : PRODUCTION_SYSPLEX
084    LAST UPDATE   : 10/20/05 14:42:02
085   DNS RR LABEL   : ZFTP
086    DNS RR STATUS : PRESENT
087    TTL           : 60
088    CLASS         : IN
089    TYPE          : A
090    RDATA         : 10.1.1.22
091    GWM LABEL     : Z/OS_LBA_ADVISOR
092    GROUP LABEL   : FTP_GROUP
093    LAST UPDATE   : 10/20/05 14:42:02
094   DNS RR LABEL   : ZFTP
095    DNS RR STATUS : PRESENT
096    TTL           : 60
097    CLASS         : IN
098    TYPE          : A
099    RDATA         : 10.1.1.1
100    GWM LABEL     : Z/OS_LBA_ADVISOR
101    GROUP LABEL   : FTP_GROUP
102    LAST UPDATE   : 10/20/05 14:42:02
103   DNS RR LABEL   : ZTELNET
104    DNS RR STATUS : PRESENT
105    TTL           : 60
```

```
106    CLASS         : IN
107    TYPE          : A
108    RDATA         : 10.1.1.22
109    GWM LABEL     : Z/OS_LBA_ADVISOR
110    GROUP LABEL   : TN3270_GROUP
111    LAST UPDATE   : 10/20/05 14:42:02
112   DNS RR LABEL   : ZTELNET
113    DNS RR STATUS : PRESENT
114    TTL           : 60
115    CLASS         : IN
116    TYPE          : A
117    RDATA         : 10.1.1.1
118    GWM LABEL     : Z/OS_LBA_ADVISOR
119    GROUP LABEL   : TN3270_GROUP
120    LAST UPDATE   : 10/20/05 14:42:02
121   DNS RR LABEL   : TELNETPRIMARY.ZTELNET
122    DNS RR STATUS : PRESENT
123    TTL           : 60
124    CLASS         : IN
125    TYPE          : A
126    RDATA         : 10.1.1.22
127    GWM LABEL     : Z/OS_LBA_ADVISOR
128    GROUP LABEL   : TN3270_GROUP
129    LAST UPDATE   : 10/20/05 14:42:02
130   DNS RR LABEL   : TELNETSECONDARY.ZTELNET
131    DNS RR STATUS : PRESENT
132    TTL           : 60
133    CLASS         : IN
134    TYPE          : A
135    RDATA         : 10.1.1.1
136    GWM LABEL     : Z/OS_LBA_ADVISOR
137    GROUP LABEL   : TN3270_GROUP
138    LAST UPDATE   : 10/20/05 14:42:02
139 1 OF 1 RECORDS DISPLAYED
```

The following comments refer to the previous display:

- Line 12:

  The TSIG FLAGS field on line 12 indicates which type of TSIG keys are being used, if any. In this example, the update_key and transfer_key keywords were not specified for the zone under this DNS. If they had been specified, an indicator or indicators would be displayed on this line showing which types of TSIG keys are being used.

- Lines 13, 49, 67, 85, 103, 121, and 130:

  The DNS RR LABEL fields indicate the portion of the DNS name of the resource record before the domain suffix is appended. Some unique examples are on lines 13, 49, 67, 85, 103, 121, and 130. Some of these labels can be compound, like TELNETPRIMARY.ZTELNET and TELNETSECONDARY.ZTELNET on lines 121 and 130. Compound labels are created for named members of server groups. The two named members were configured in a server_group statement on lines 87 and 97 of the example ADNR configuration file. The label PRODPLEX on line 13 represents the resource records for the virtual unnamed member of the host_group statement on line 26 of the example ADNR configuration file. The labels SYSA on line 49 and SYSB on line 67 represent the named members of the host group on line 36 and 45 of the example ADNR configuration file. The label ZFTP on line 85 represents the unnamed member of the server group on line 120 of the example ADNR configuration file. The label ZTELNET on line 103 represents the virtual unnamed member of the server group on line 108 of the example ADNR configuration file. The label TELNETPRIMARY.ZTELNET on line 121 represents the named member on line 87 of the example ADNR configuration file, and the label TELNETSECONDARY.ZTELNET on line 130 represents the named member on line 97 of the example ADNR configuration file.

- Line 14:

  The fields labeled DNS RR STATUS (for example, line 14) indicate whether or not the resource record is present in the name server. Resource records representing available resources in the sysplex have a value of PRESENT, while resource records representing unavailable resources in the sysplex have a value of NOT_PRESENT.

- Line 15:

The fields labeled TTL (for example, line 15) indicate the time to live value associated with the resource record. For more information, see "Near real-time availability information of sysplex resources" on page 1209.

- Line 16:

  The fields labeled CLASS (for example, line 16) indicate the resource record class. All ADNR-managed resource records have a class value of IN.

- Line 17:

  The fields labeled TYPE (for example, line 17) indicate the resource record type. ADNR-managed resource records are either A for IPv4 addresses or AAAA for IPv6 addresses.

- Line 18:

  The fields labeled RDATA (for example, line 18) contain the IPv4 or IPv6 address that is in the resource record.

- Line 19:

  The fields labeled GWM LABEL (for example, line 19) match the label of the gwm statement on line 20 of the example ADNR configuration file.

- Line 20:

  The fields labeled GROUP LABEL (for example, line 20) indicate the group definition that is responsible for the creation of the resource record. This makes it possible to correlate this resource record with the MODIFY ADNR,DISPLAY,GWM,GROUPS,GROUPID=PRODUCTION_SYSPLEX output to see the GWM data that is related to this resource record.

- Line 21:

  The fields labeled LAST UPDATE (for example, line 21) indicate the most recent time that the status of the resource record has changed from PRESENT to NOT_PRESENT, or from NOT_PRESENT to PRESENT.

- Line 139:

  The number of records displayed on line 139 indicates the number of zones actually displayed. When the number of zones defined exceeds the MAX parameter value (or the default of 100) of the MODIFY ADNR,DISPLAY command, the difference between the numbers on this line indicates the number of zones that were omitted from the display as a result of the MAX parameter value.

Chapter 23. Simple Network Management Protocol

This topic describes how to configure:

- Simple Network Management Protocol (SNMP) agent (osnmpd)
- z/OS UNIX **snmp** or **osnmp** command
- NetView SNMP command
- SNMP subagents
- Open Systems Adapter (OSA) support
- Trap forwarder daemon

Before you configure, read "Understanding search orders of configuration information" on page 13. It covers important information about data set naming and search sequences.

SNMP overview

SNMP is a set of protocols that describes management data and the protocols for exchanging that data between heterogeneous systems. The protocols include both the description of the management data, defined in the Management Information Base (MIB), and the operations for exchanging or changing that information. By implementing common protocols, management data can be exchanged between different platforms with relative ease.

SNMP defines an architecture that consists of:

- Network management applications
- Network management agents and subagents
- Network elements, such as hosts and gateways

The SNMP network management application can ask agents for specific information about network elements. Conversely, agents can tell the network management application when something happens to one or more network elements. The protocol used between the network management application and agents is SNMP. The transport protocol for SNMP requests is the User Datagram Protocol (UDP).

SNMP defines both the network management data and the ways in which the data is retrieved or changed by the network management application. Examples of network management data include device definitions, counts of packets received at the IP layer, TCP connection data, and so forth. The information about the network elements is stored in the Management Information Base (MIB), which is supported by the SNMP agent and its subagents. A MIB variable (or MIB object) is a specific instance of data in a collection of objects related to a common management area. The collection is called a MIB module. Each MIB object is identified by an object identifier (OID), which is a dotted-decimal value, and by a textual name. For example, the sysUpTime MIB object, which indicates how long ago the SNMP agent initialized, is defined as follows:

```
Textual name     Object identifier (OID)
sysUpTime        1.3.6.1.2.1.1.3
```

The z/OS Communications Server Network Management agent and subagents, also called SNMP agent and SNMP subagents, support many standard (RFC-based) MIB modules. The SNMP TCP/IP subagent also supports enterprise-specific MIB modules. For information on MIB modules supported by the z/OS Communications Server SNMP agent and subagents, see the SNMP agent capabilities statement, shipped as file /usr/lpp/tcpip/samples/mvstcpip.caps. Additionally, enterprise-specific MIBs are documented in the /usr/lpp/tcpip/samples directory. See "TCP/IP subagent" on page 1240 for more information on enterprise-specific MIB modules supported by the TCP/IP subagent. For a complete list of MIB objects supported by the SNMP agent and subagents shipped with z/OS Communications Server, see z/OS Communications Server: IP System Administrator's Commands.

© Copyright IBM Corp. 2000, 2022

Network management application

The stations that monitor network elements run network management applications. In z/OS Communications Server, the z/OS UNIX **snmp** (or **osnmp**) command provides SNMP network management from the z/OS UNIX shell. The NetView SNMP command provides network management from the NetView command line. The z/OS UNIX **snmp** and **osnmp** commands perform exactly the same function. Both commands can be used to retrieve or change data from SNMP and monitor for asynchronous events known as notifications. An unconfirmed notification is called a trap. A confirmed notification is called an inform.

For information about the syntax and use of the **snmp** and NetView SNMP commands, see z/OS Communications Server: IP System Administrator's Commands.

SNMP protocols

This topic provides an overview of the different SNMP protocols and their capabilities. For more detailed information on the security models for each of these SNMP protocols, see "Overview of SNMP security models" on page 1239.

SNMPv1

SNMPv1, standardized in 1988, is the first and most commonly used version. It became very popular and is probably the most widely deployed of the SNMP generations today. The SNMPv1 standards define support for SNMP GET, GETNEXT, and SET operations, and for asynchronous event notifications called TRAPs. However, SNMPv1 does not support some later MIB object types, such as 64-bit counters. SNMPv1 also uses community-based security, which is not very secure.

SNMPv2

The SNMPv2 protocol standards made several attempts to address the security issues associated with the SNMPv1 protocol, with the party-based security model SNMPv2p, the user-based security model SNMPv2u, and the community-based security model SNMPv2c.

Although these attempts were not successful in addressing the major security issues, SNMPv2 did provide several improvements over SNMPv1, especially in the area of data retrieval with the support of SNMP GETBULK operations, and SNMPv2 continued providing community-based security with SNMPv2c.

SNMPv3

SNMPv3 was designed in the late 1990s, and in December of 2002 become a standard. It is defined in RFCs 3410 through 3415 [see Appendix G, "Related protocol specifications," on page 1409]. SNMPv3 uses the basic SNMP management system and operations of SNMPv1 and SNMPv2, but adds an entirely new security architecture. The SNMPv3 architecture is modularized so that portions of it can be enhanced over time without requiring that the entire architecture to be replaced. SNMPv3 defines a framework that, among other things, includes the following models:

- Message processing model (SNMPv3)
- User-based security model
- View-based access control model

The framework is structured so that multiple models can be supported concurrently and replaced over time. For example, although there is a new message format for SNMPv3, messages that are created with the SNMPv1 and SNMPv2c formats are still supported. Similarly, the user-based security model can be supported concurrently with previously used community-based security models. In addition, SNMPv3 added other key updates to the protocol, such as the following updates:

- Improved notification support

  A new notification type, INFORM, is like a TRAP that requires an acknowledgement. If the acknowledgment is not received, the INFORM is resent.

- Trap filtering

  With SNMPv3, a TRAP can also be filtered at the sender.
- Dynamic configuration

  The SNMP agent can be dynamically configured using MIB modules defined in RFC 3584 and RFCs 3411 through 3415.

SNMP agent

In z/OS Communications Server, the SNMP agent is a z/OS UNIX application. It supports a maximum SNMP response packet size of 65535 bytes, and supports the following SNMP protocols:

- SNMPv1
- SNMPv2c
- SNMPv3

SNMPv2c offers protocol enhancements such as the GETBULK operation. SNMPv3 provides a network management framework that allows the use of user-based security in addition to, or instead of, the community-based security supported in SNMPv1 and SNMPv2c. The view-based access control model supported in SNMPv3 allows granular access control for MIB objects with either the user-based or community-based security models. SNMPv3 also enables dynamic changes to the SNMP agent configuration.

Overview of SNMP security models

SNMP security has evolved from community-based security to a modularized architecture that provides message security and access control.

SNMPv1 and SNMPv2c

The security model used by SNMPv1 and SNMPv2c is the community-based security model. In this model, an SNMP community is made up of an SNMP agent along with SNMP manager entities. Managers typically request management data from the agents. Each SNMP community is represented using an octet string called the community name. When the manager communicates with the agents, it uses the community name as a password to get access to the management resources. Because the community name is sent in every packet in clear view, these communications are not secure.

The access control mechanism provided by SNMPv1 is very simple. A community is allowed access to read or write objects in a management information base (MIB) tree. In most implementations, a community has access to all of the objects in the MIB, and you cannot restrict the access to a particular part of the MIB tree.

SNMPv3

SNMPv3 addresses the basic lack of security inherent in the previous SNMP versions by providing message security and access control. For message security, it introduces the User-Based Security Model (USM), which provides for authentication and privacy. Additionally, access control is provided with View-Based Access Control Model (VACM). Both USM and VACM provide for secure communications when you use SNMPv3.

User-Based Security Model (USM)
This model was designed to provide message security. USM supports both authentication (data integrity, data authentication) and privacy (protection against disclosure of message payload). For authentication, the protocols supported are HMAC-MD5 and HMAC-SHA. For privacy, CBC-DES 56-bit and AES 128-bit CFB mode are the supported symmetric encryption protocols. If you use the AES protocol, the z/OS Integrated Cryptographic Service Facility (ICSF) must be active. For detailed information about configuring ICSF, see z/OS Cryptographic Services ICSF Administrator's Guide.

View-Based Access Control Model (VACM)
VACM is used to provide access control. With VACM, users are defined to groups that are allowed to access different views or parts of the management data (MIB objects), depending on defined data access privileges.

SNMP subagents

A subagent extends the set of MIB variables supported by an SNMP agent. z/OS Communications Server supports the following subagents:

- TCP/IP subagent
- OMPROUTE subagent
- TN3270E Telnet subagent
- Network SLAPM2 subagent

For a complete list of MIB objects, see z/OS Communications Server: IP System Administrator's Commands.

The OSA-Express Direct subagent is also shipped with the z/OS Communications Server product, but is supported by the System z OSA Support Group.

TCP/IP subagent

The TCP/IP subagent in z/OS Communications Server is a z/OS UNIX application that runs in its own task in the TCP/IP address space. This subagent supports many standard (RFC-based) MIB objects. In addition, it supports MIB objects in the following enterprise-specific MIB modules:

- The IBM 3172 enterprise-specific MIB.
- The IBM MVS TCP/IP enterprise-specific MIB. This MIB defines objects to extend standard MIB tables and supports retrieval and change of TCP/IP address space configuration parameters.

For a more detailed list of all the data supported by the TCP/IP subagent, see z/OS Communications Server: IP System Administrator's Commands.

Both the TCP/IP subagent and the OSA-Express Direct subagent support management data for OSA-Express features. You should use the OSA-Express Direct subagent to obtain this management data, for the following reasons:

- Because the OSA-Express Direct subagent communicates directly with the OSA-Express features, it does not require the OSA/SF and IOASNMP applications.
- The OSA-Express Direct subagent provides more OSA management data than the TCP/IP subagent.
- Because of the support provided by the OSA-Express Direct subagent, there will be no new enhancements to the OSA management data provided by the TCP/IP subagent, and support for this data will eventually be removed in a future release.

If you are using the TCP/IP subagent's OSA management data support, and decide to switch to using the OSA-Express Direct subagent instead, you no longer need to start the OSA/SF address space and the OSA IOASNMP application. For more information about this subagent, see “OSA-Express Direct subagent” on page 1241. For information on the TCP/IP subagent's dependency on OSA/SF and IOASNMP, see “Step 4: Configure the Open Systems Adapter support” on page 1261.

The TCP/IP subagent support evolves with each release. New MIB objects might be supported, and, occasionally, old ones might be removed for functions that are no longer relevant. This is particularly true for the MIB objects in the IBM MVS TCP/IP enterprise-specific MIB. This MIB's definition is shipped as file /usr/lpp/tcpip/samples/mvstcpip.mi2. For details of the changes for each release, see the REVISION sections of this MIB module file.

The TCP/IP subagent in z/OS Communications Server provides SET support, enabling remote configuration of some TCP/IP address space parameters. The TCP/IP subagent is configured and controlled by the SACONFIG statement in the PROFILE.TCPIP data set. Systems where SNMP support is not required can disable the subagent and save system resources.

OMPROUTE subagent

The OMPROUTE subagent implements the Open Shortest Path First (OSPF) MIB variable containing OSPF protocol and state information.

The OMPROUTE subagent supports selected MIB objects defined in RFC 1850.

For a detailed description of the OMPROUTE subagent, see z/OS Communications Server: IP Configuration Reference.

TN3270E Telnet subagent

The SNMP TN3270E Telnet subagent provides Telnet transaction data for monitored Telnet connections using the SNMP protocol. For more information about configuring the TN3270E Telnet subagent, see the TNSACONFIG statement in z/OS Communications Server: IP Configuration Reference.

Network SLAPM2 subagent

The z/OS Communications Server Network SLAPM2 subagent (nslapm2) enables network administrators to retrieve data to determine whether the current set of Network SLAPM2 policy definitions are performing as needed or whether adjustments need to be made. The Network SLAPM2 subagent supports the Network Service Level Agreement Performance Monitor (NETWORK-SLAPM2) MIB. For more information about the Network SLAPM2 MIB, see usr/lpp/tcpip/samples/slapm2.mi2.

For a detailed description of the nslapm2 subagent, see the information on Policy Agent and policy applications in z/OS Communications Server: IP Configuration Reference.

OSA-Express Direct subagent

The OSA-Express Direct subagent supports management data for OSA-Express features. The OSA-Express Direct subagent is shipped with the z/OS Communications Server product, but is supported by the System z OSA Support Group. The MVS started procedure name of this subagent is IOBSNMP. For a complete understanding of the management data provided by the OSA-Express Direct subagent, and information on configuring the subagent, see Open Systems Adapter-Express Customer's Guide and Reference.

Tip: The *-c* parameter in the IOBSNMP MVS started procedure, specifies the SNMP community name value for connections between the OSA-Express Direct subagent and the SNMP agent, OSNMPD. If you want to avoid externalizing this value for security reasons, you can use the MVS JCL PARMDD parameter on the EXEC statement to pass all the parameters to the subagent in a data set. For example, you could set up your IOBSNMP started procedure as follows:

```
//IOBSNMP  PROC
//*
//* Using PARMINDD for the PARMs that were on the PROC line
//*
//IOBSNMP  EXEC PGM=IOBSNMP,TIME=1440,REGION=4096K,DYNAMNBR=5,
//         PARMDD=PARMINDD
//SYSPRINT DD  SYSOUT=*
//SYSUDUMP DD  SYSOUT=*
//PARMINDD DD  DISP=SHR,DSN=SYS1.IOBSNMP.PARMS

The SYS1.IOBSNMP.PARMS data set could contain:
  -c my_community_name -d 0 -s TCPIP
```

For more information on the PARMDD parameter, see z/OS MVS JCL Reference.

The TCP/IP subagent also supports management data for OSA-Express features, but you should use the OSA-Express Direct subagent to obtain this management data for the following reasons:

- The OSA-Express Direct subagent communicates directly with the OSA-Express features and does not require the OSA/SF and IOASNMP applications.
- The OSA-Express Direct subagent provides more OSA management data than the TCP/IP subagent.

- The TCP/IP subagent does not support management data for OSA-Express3 or later features.
- Because of the support provided by the OSA-Express Direct subagent, there will be no new enhancements to the OSA management data provided by the TCP/IP subagent, and support for this data will eventually be removed in a future release.

If you are using the TCP/IP subagent's OSA management data support, and decide to switch to using the OSA-Express Direct subagent instead, you no longer need to start the OSA/SF address space and the OSA IOASNMP application.

Key generation commands

The **pwtokey** and **pwchange** commands are provided to enable generation and change of keys used for authentication and encryption with SNMPv3.

For more information about **pwtokey**, see z/OS Communications Server: IP Configuration Reference. For information about **pwchange**, see z/OS Communications Server: IP System Administrator's Commands.

Distributed Protocol Interface

The DPI is an application interface used by the SNMP agent to communicate with subagents. With DPI, you can dynamically add, delete or replace management variables supported by the SNMP agent and its subagents. z/OS Communications Server provides DPI V2.0 for z/OS UNIX C socket users and DPI V1.1 for traditional C socket users. DPI V2.0 provides additional function, making it easier to write subagents and simplifying the task of developing and administering your application.

The SNMP agent also listens on a TCP ephemeral port number (>1023) that is used for subagent communication via the DPI. You cannot control this port number because of the way ephemeral ports are given out. When the SNMP agent asks for an ephemeral port, it gets the next one available. See RFC 1228 for more information.

For more information about the DPI, see z/OS Communications Server: IP Programmer's Guide and Reference.

Trap forwarder daemon

The trap forwarder daemon enables multiple SNMP managers to receive SNMP traps from the same TCP/IP stack. The trap forwarder daemon listens for traps on a port, usually the well-known port 162, and forwards the traps to all configured managers.

For more information about the trap forwarder daemon, see z/OS Communications Server: IP Configuration Reference.

Processing an SNMP request

Figure 151 on page 1242 illustrates the flow of processing an SNMP request when the request is for data supported by the TCP/IP subagent.

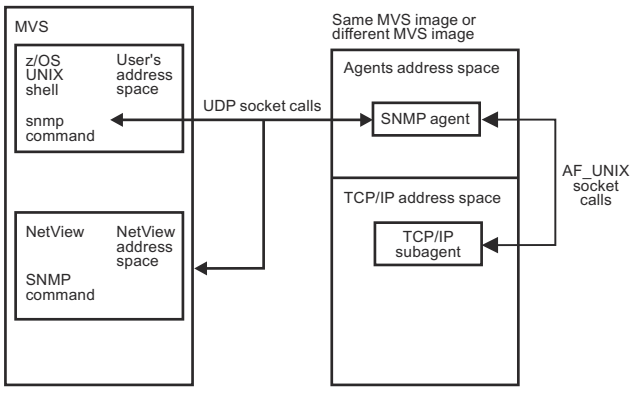


Figure 151. Overview of SNMP support

The following list describes the sequence of events that occur from the time you issue an SNMP command until you receive the response:

1. The user issues a NetView SNMP command or a z/OS UNIX **snmp** command.
2. The command processor validates and encodes the request in a Protocol Data Unit (PDU), and sends it in a UDP packet to the SNMP agent. The SNMP agent can be on the same MVS image or on a different MVS image.
3. The SNMP agent validates the request and, if necessary, sends it to an SNMP subagent. Requests for agent-oriented objects are handled by the agent and all others are handled by a subagent. To determine which objects are handled by the agent and which objects are handled by a subagent, see the *Management Information Base (MIB) objects* appendix in z/OS Communications Server: IP System Administrator's Commands.
4. The agent sends the response to the originator of the request. The command processor displays the response.

Tip: Although not shown in Figure 151 on page 1242, other subagents that are shipped as part of z/OS Communications Server, such as the OMPROUTE subagent, the Network SLAPM2 subagent, and the TN3270E Telnet subagent, also communicate with the SNMP agent using AF_UNIX socket calls from their own address spaces.

The SNMP agent and the SNMP subagents record trace information through the z/OS UNIX syslog daemon (syslogd) using the *daemon* facility. For more information about syslogd and specifying the daemon facility in the /etc/syslog.conf configuration file, see "Logging of system messages" on page 30.

Deciding on SNMP security needs

The SNMP agent supports SNMPv1, SNMPv2c, and SNMPv3 security.

Community-based security

SNMPv1 and SNMPv2c are community-based security, where a community name (or password) is passed with a request. If the community name is recognized as one that can be used by the IP address from which the request originates, the SNMP agent processes the request.

User-based security

SNMPv3 provides a more powerful and flexible framework for message security and access control. Message security involves providing:

- Data integrity checking, to ensure that the data was not altered in transit
- Data origin verification, to ensure that the request or response originates from the source from which it claims to have come
- Message timeliness checking and, optionally, data confidentiality, to protect against eavesdropping

Access control is the ability to control exactly what data an individual user can read or write.

The SNMPv3 architecture introduces the User-Based Security Model (USM) for message security and the View-Based Access Control Model (VACM) for access control. The architecture supports the concurrent use of different security, access control, and message processing models. For example, community-based security can be used concurrently with USM.

USM uses the concept of a user for which security parameters (levels of security, authentication and privacy protocols, and keys) are configured at both the agent and the manager. Messages sent using USM are better protected than messages sent with community-based security, where passwords are sent in the clear and displayed in traces. With USM, messages exchanged between the manager and the agent have data integrity checking and data origin authentication. Message delays and message replays (beyond what happens normally due to a connectionless transport protocol) are protected against with the use of time indicators and request IDs. Data confidentiality, or encryption, is also available.

The use of VACM involves defining collections of data (called views), groups of users of the data, and access statements that define which views a particular group of users can use for reading, writing, or receipt in a notification.

The SNMP agent can be configured to use USM and VACM by specifying SNMPD.CONF information. SNMPv3 also introduces the ability to dynamically configure the SNMP agent using SNMP SET commands against the MIB objects that represent the agent's configuration. These MIB objects are defined in RFC 3584 and RFC 3411 through 3415. This dynamic configuration support enables addition, deletion, and modification of configuration entries either locally or remotely. Remote modification of user keys can be especially useful.

Decide on your security needs - community-based or user-based

If you are satisfied with the security of your existing configuration, you can continue to use community-based security with no migration. If you would like to take advantage of USM or VACM, or if you have some SNMP managers that use SNMPv3, you will need to migrate your configuration. Note that USM can be used only when both the SNMP agent and the manager requesting the data support USM, as the z/OS Communications Server SNMP agent and the **snmp** command do. VACM can be used even for community-based requests, but doing so requires migration of existing community name and trap destination definitions in PW.SRC and SNMPTRAP.DEST to SNMPD.CONF.

Even if your managers continue to be community-based, there are important advantages to migrating your PW.SRC information to SNMPD.CONF format:

- Enables users to make use of the access control mechanism provided with SNMPv3 with community-based security.
- Provides the ability to dynamically configure the z/OS SNMP agent using MIBs.
- Provides a way of easing into SNMPv3 user-based security.
- Does not require any changes to the manager configuration.

The following tables list the advantages and disadvantages of using each type of security.

Table 70. SNMPv1/SNMPv2c advantages and SNMPv3 disadvantages

SNMPv1/SNMPv2c advantages	**SNMPv3 disadvantages**
Widely implemented on many platforms.	Not yet implemented on many platforms.
Easy to configure.	More robust configuration options.

Table 71. SNMPv1/SNMPv2c disadvantages and SNMPv3 advantages

SNMPv1/SNMPv2c disadvantages	**SNMPv3 advantages**
Legacy standards-based administrative model.	New standards-based administrative model.
SNMPv1 and SNMPv2c allow particular IP addresses to access all data or no data.	SNMPv3 allows a particular user to access particular data.
Not very robust (password sent in PDU).	Robust (data integrity and data origin authentication).
Any user that can read data can also change the data (for objects defined as read-write).	The ability to change data can be limited to specific users.
No data confidentiality.	Encryption available.
Configuration changes require restarting of SNMP agent.	Configuration changes for USM and VACM can be made dynamically, either locally or remotely.

For more information about security, see “Creating user keys” on page 1251.

Step 1: Configure the SNMP agent

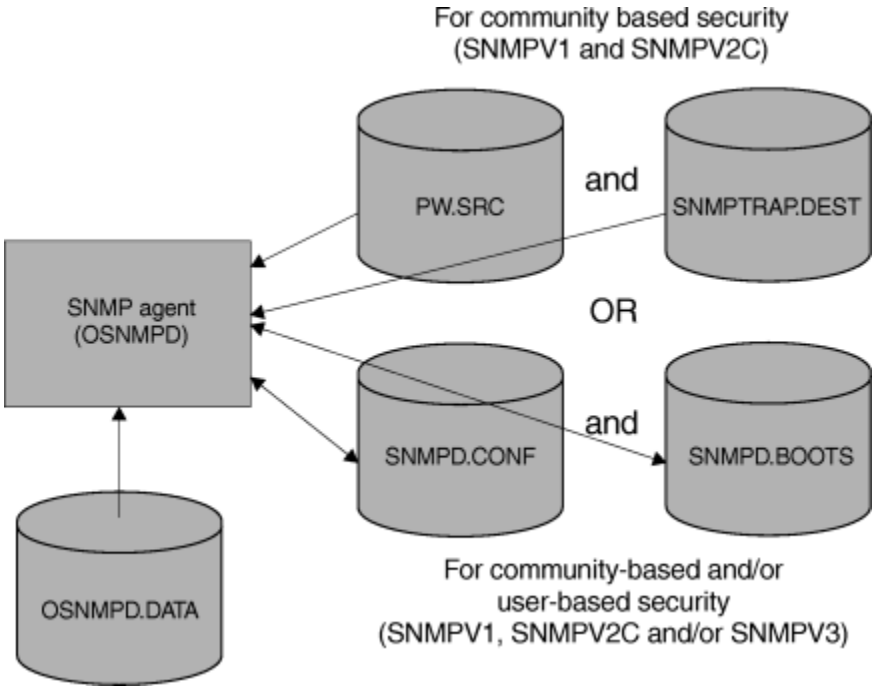


Figure 152. Configuration files for SNMP agent

Configure the SNMP agent (OSNMPD) based upon your security need. The SNMP agent accepts both SNMPv1 and SNMPv2c requests for community-based security. The SNMP agent can be configured to also use the User-based Security Model and the View-based Access Control Model. You should assign the SNMP agent and all the SNMP subagents to the same WLM service class so that they all have the same dispatching priority. Timeouts can occur if the SNMP subagents are set to a lower dispatching priority than the SNMP agent.

To configure the SNMP agent, perform the following tasks:

- "Provide TCP/IP profile statements" on page 1245
- Depending upon whether you want to use USM and VACM, see one of the following topics:
  - If you are using community-based security and do not need USM or VACM, see "Provide community-based security and notification destination information" on page 1246.
  - If you want the flexibility of using USM or VACM or community-based security, see "Provide community-based and user-based security and notification destination information" on page 1249.
- "Provide MIB object configuration information" on page 1253
- See z/OS Communications Server: IP Configuration Reference for more information about OSNMPD parameters.

Provide TCP/IP profile statements

Update the following configuration statements in *hlq*.PROFILE.TCPIP:

AUTOLOG
PORT

There are two primary TCP/IP ports used by the SNMP agent, one for receiving incoming requests and one for sending traps to managers.

The default port used by the SNMP agent to receive incoming requests is 161. If you want the agent to use port 161 for this purpose and want to ensure that no other application uses this port, you must specify the following PORT statement in your profile data set:

```
PORT
  161 UDP OSNMPD              ; SNMP Agent port for SNMP requests
```

If the agent will be started from the z/OS shell, reserve the port instead for z/OS UNIX by typing *OMVS* instead of *OSNMPD*.

If you want to define a port other than 161 for SNMP requests, take the following steps:

1. Start the agent with the `-p` parameter.
2. Configure management applications to use the new port:
   - For the **snmp** command, make an entry in the OSNMP.CONF file with the correct port number. For details on creating this entry, see the description for *targetAgent* in the OSNMP.CONF statement in the z/OS Communications Server: IP Configuration Reference.
   - Where supported, configure other management applications to use the new port.
3. Configure subagents to use the new port:
   a. Specify the port number to use on the SACONFIG profile statement for the TCP/IP subagent.
   b. Specify the port number to use on the ROUTESA_CONFIG profile statement for the OMPROUTE subagent.
   c. Specify the port number to use on the `-p` parameter when starting the Network SLAPM2 subagent.
   d. Specify the port number to use on the TNSACONFIG profile statement for the TN3270E Telnet subagent.
   e. If you are using DPI subagents other than those supplied with z/OS Communications Server, set the SNMP_PORT environment variable to enable user-written subagents to connect to the agent.

The SNMP agent uses port 162, by default, for sending traps to the managers specified in SNMPTRAP.DEST or SNMPD.CONF file. Port 162 should be reserved for the management application primarily responsible for trap processing. If your environment requires multiple management applications at the same IP address to receive traps, consider using the Trap Forwarder Daemon. See “Step 5: Configure the trap forwarder daemon” on page 1265 for more details. If the SNMP query engine is typically used for processing traps and other applications, such as snmp, that are only occasionally used, the following port reservations are recommended.

```
PORT
  162 UDP SNMPQE  ; SNMPQuery Engine
```

You must also reserve additional ports for use by the **snmp** command by specifying

```
nnnnn UDP OMVS
```

where *nnnnn* is a number in the range 0–65535 and *nnnnn* is used as the `-p` parameter value on the **snmp trap** command.

If you want the SNMPQE and OSNMPD address spaces to be started automatically when the TCPIP address space is started, then include SNMPQE and OSNMPD in the AUTOLOG statement:

```
AUTOLOG
  SNMPQE                      ; SNMP Query Engine
  OSNMPD                      ; SNMP Agent
ENDAUTOLOG
```

Provide community-based security and notification destination information

If you are using only community-based security without the view-based access control model, see the following subtopics to configure the security and trap destinations.

Provide community name information

SNMP agents are accessed by remote network management stations and by SNMP subagents. To allow network management stations to send inquiries to the SNMP agent, and SNMP subagents to connect to the SNMP agent, you can provide PW.SRC information that defines a list of community names and IP addresses that can use these community names. The community name operates as a password when accessing objects on, or connecting to, a destination SNMP agent. The subagents pass the community name to the agent on the connect request.

Guideline: Because the community name of public is a well-known name, it should not be configured to the agent due to security considerations. IBM Health Checker for z/OS can be used to check whether the community name of public has been configured to the SNMP agent. For more details about IBM Health Checker, see IBM Health Checker for z/OS: User's Guide.

All of the z/OS Communications Server SNMP subagents connect to the agent using the IPv4 primary interface IP address of the stack with which the subagent is associated, and a community name. As long as SOURCEVIPA is not in effect on the IPCONFIG profile statement, this IP address is the source IP address that the agent uses, along with the community name, to verify the subagent's authority to connect to the SNMP agent. The IPv4 primary interface IP address is either the first IP address in the HOME list or the IP address specified on a PRIMARYINTERFACE TCP/IP profile statement. If SOURCEVIPA is in effect, the IP address used by the agent to verify the subagent's authority is the virtual IP address associated with the IPv4 primary interface IP address. For information on determining which virtual IPv4 address is associated with a physical IPv4 address, see the HOME statement in z/OS Communications Server: IP Configuration Reference. Check the Netstat HOME/-h output to verify the IPv4 primary interface address of the stack.

The PW.SRC information is optional. If no PW.SRC information is found and no community name is specified for the `-c` parameter at agent invocation, then the SNMP agent will accept requests with a community name of 'public' from any IP address. If a PW.SRC file exists, but is empty, and if no community name is specified on the `-c` parameter at the agent invocation, then no requests will be accepted by the agent.

Note: Verify that there is no SNMPD.CONF file because this file can be used only with SNMPv3. If an SNMPD.CONF file is found, the PW.SRC file will not be used.

If creating a data set, you can specify a sequential data set with the following attributes: RECFM=FB, LRECL=80, and BLKSZ=3120. Other data set attributes might also work, depending on your installation parameters.

PW.SRC example

The PW.SRC statements could be specified as follows:

```
passwd1 9.0.0.0      255.0.0.0
passwd2 129.34.81.22 255.255.255.255
IPv6passwd3 12ab::0 16
IPv6passwd4 39B3::F430:03EE 128
```

The PW.SRC statements specify community names and hosts that can use each community name. The format of a statement is:

community_namedesired_networksnmp_mask

See z/OS Communications Server: IP Configuration Reference for more information about syntax.

The community name of an incoming SNMP request is compared to the known community names. If a match is found, then the IP address of the incoming request is logically ANDed with the *snmp_mask* of the PW.SRC statement. The result of the logical ANDing process is compared with the *desired_network*. If they match, the request is accepted.

In the case of a password definition to be used by an IPv6 address or range of IPv6 addresses, the *snmp_mask* can be specified as a prefix value. The prefix specifies the number of bits to be used to construct an IPv6 address mask.

In the preceding example, if a request for *community_name* passwd1 is received from the IP address 9.34.22.122, IP address 9.34.22.122 is ANDed with 255.0.0.0. The result is 9.0.0.0, which equals the specified *desired_network* for passwd1, so this request is accepted. In passwd2, if the *community_names* match, only requests from host 129.34.81.22 are accepted. The password IPv6passwd3 can be used by any IPv6 address that starts with 12ab.

If the *community_name* values do not match, or the IP address ANDed with the *snmp_mask* does not match, an AUTHENTICATION_FAILURE trap is sent if both of the following conditions are true:

- A destination entry exists in SNMPTRAP.DEST.
- Authentication failure traps have been enabled. These traps are enabled by either setting MIB object snmpEnableAuthenTraps.0 to 1, or specifying the following statement in the OSNMPD.DATA configuration information:

```
snmpEnableAuthenTraps 1
```

A *desired_network* and *snmp_mask* of all zeros allows anyone with the correct *community_name* to make requests. However, the passwords for IPv4 addresses and the passwords for IPv6 addresses are stored and handled separately. Defining a password for use by both IPv4 and IPv6 addresses requires two entries in PW.SRC. Likewise, defining a password to be used by all addresses (both IPv4 and IPv6) requires two entries as follows:

```
passwd5 0.0.0.0 0.0.0.0
passwd5 0::0 0
```

Note: By default, the SNMP agent and the **snmp** command send packets such that a VIPA address will be used as the originating address in the packet, if SOURCEVIPA is configured. This is a change introduced in V2R10; previously, the SNMP agent and the **snmp** command set a socket option to cause the physical interface addresses to be used as the originating addresses on packets they sent. That meant the PW.SRC file had to contain all of the possible physical interface addresses that might be used, rather than a smaller number of VIPA addresses. A customer can override this change in behavior, if wanted. This can be done for the SNMP agent by invoking it with the -a option. Similarly, you can do the same for the UNIX **snmp** command by either invoking it from the command line with the -a option, or by coding NOSVIPA in the command's OSNMP.CONF configuration file.

Provide trap destination information

Traps are unsolicited messages that are sent by an SNMP agent to an SNMP network management station. An SNMP trap contains information about a significant network event. The management application running at the management station interprets the trap information sent by the SNMP agent.

Note: When the SNMP agent starts, it retrieves an IP address for itself. If it retrieves an IPv6 colon-hexadecimal address, when it sends traps the source IP address in each trap will be 0.0.0.0.

For a detailed description of the SNMP trap types provided by z/OS CS, see z/OS Communications Server: IP System Administrator's Commands.

The SNMP agent Distributed Protocol Interface allows subagents other than those shipped with z/OS Communications Server (which might be running on another host) to generate SNMP traps. This can allow for support of other types of traps. For more information about SNMP DPI, see the z/OS Communications Server: IP Programmer's Guide and Reference.

To use traps, you must provide SNMPTRAP.DEST information defining a list of managers to which traps are sent. The SNMPTRAP.DEST information is optional. If no trap destination file is found, then the SNMP agent sends traps to the IP address of the SNMP agent and issues a warning message indicating that defaults are in effect. If a trap destination file exists, but is empty, no traps are sent.

Guideline: If you use SNMPTRAP.DEST to configure trap information, the agent uses the hardcoded community name of public in the outbound traps. Because the community name of public is a well-known name, it should not be used in SNMP traps due to security considerations. To configure specific community names for trap destinations, you must convert your SNMPTRAP.DEST information to a SNMPD.CONF configuration file format. For more information about how to accomplish this conversion,

see Migrating the PW.SRC file and SNMPTRAP.DEST file to the SNMPD.CONF file in z/OS Communications Server: IP Configuration Reference.

Note: Verify that there is no SNMPD.CONF file. If an SNMPD.CONF file is found, the SNMPTRAP.DEST file will not be used.

If creating a data set, you can specify a sequential data set with the following attributes: RECFM=FB, LRECL=80, and BLKSZ=3120. Other data set attributes might also work, depending on your installation parameters.

SNMPTRAP.DEST example

The SNMPTRAP.DEST statements could be specified as follows:

```
# SNMP Trap Destination information
124.34.216.1 UDP
39B3::F430:03EE UDP
MVSSYS2      UDP
```

See z/OS Communications Server: IP Configuration Reference for more information about syntax.

Provide community-based and user-based security and notification destination information

If you want to use user-based security, either concurrently with or instead of community-based security, you must configure security definitions and notification destinations. To allow SNMP subagents to connect to the SNMP agent using user-based security, you must configure community-based security definitions. The SNMP subagents pass the community name to the agent on the connect request. The community name operates as a password when the SNMP subagents connect to the SNMP agent.

Guideline: Because the community name of public is a well-known name, it should not be configured to the agent due to security considerations. IBM Health Checker for z/OS can be used to check whether the community name of public has been configured to the SNMP agent. For more details about IBM Health Checker, see IBM Health Checker for z/OS: User's Guide.

All of the z/OS Communications Server SNMP subagents connect to the agent using the IPv4 primary interface IP address of the stack with which the subagent is associated, and a community name. As long as SOURCEVIPA is not in effect on the IPCONFIG profile statement, this IP address is the source IP address that the agent uses, along with the community name, to verify the subagent's authority to connect to the SNMP agent. The IPv4 primary interface IP address is either the first IP address in the HOME list or the IP address specified on a PRIMARYINTERFACE TCP/IP profile statement. If SOURCEVIPA is in effect, the IP address used by the agent to verify the subagent's authority is the virtual IP address associated with the IPv4 primary interface IP address. For information on determining which virtual IPv4 address is associated with a physical IPv4 address, see the HOME statement in z/OS Communications Server: IP Configuration Reference. Check the Netstat HOME/-h output to verify the IPv4 primary interface address of the stack.

SNMPv3 provides the ability to configure the agent dynamically, from either a local or remote host, and to make changes in the configuration while the SNMP agent is running. Doing SNMP agent configuration dynamically requires a good understanding of how the SNMP SET commands can be issued to create new rows or to change or delete existing rows, as well as familiarity with the SNMP engine configuration tables defined in RFCs 3584 and 3411 through 3415. For information about accessing RFCs, see Appendix G, “Related protocol specifications,” on page 1409.

As an alternative to dynamically configuring the SNMP agent, z/OS Communications Server supports a configuration file to be read at agent initialization called the SNMPD.CONF file. Dynamic configuration changes made with SNMP SET commands to the SNMP agent configuration entries will be written out to the SNMPD.CONF file, so they will continue to be in effect even after the SNMP agent is restarted.

SNMPD.CONF file

The SNMPD.CONF file defines the SNMP agent security and notification destinations. If the SNMPD.CONF file exists, the agent can support SNMPv1, SNMPv2c, and SNMPv3 requests. If no SNMPD.CONF file exists, the agent will support only SNMPv1 and SNMPv2c requests.See Coding the SNMPD.CONF entries in z/OS Communications Server: IP Configuration Reference for descriptions of the SNMPD.CONF configuration statements.

Note: If the SNMPD.CONF file is found, the PW.SRC file and the SNMPTRAP.DEST files are not used.

SNMPD.CONF dynamic configuration

If the SNMPD.CONF information is located in an MVS data set rather than a z/OS UNIX file, special considerations must be made to support dynamic configuration changes to the SNMP agent's configuration. If dynamic configuration changes are made, the file is rewritten to reflect the changes. Therefore, consider the following items when allocating the SNMPD.CONF file to an MVS data set:

- The record length (LRECL) should be 512 bytes to accommodate the longest possible entry.
- The use of a member of a partitioned data set is tolerated but not recommended. Because the file might be rewritten often, frequent compression of the partitioned data set may become necessary. In addition, locking on the file is done at the data set level, not at the member level, so other members of the partitioned data set would not be usable while the SNMP agent was running (after a dynamic configuration change had been made).

SNMPD.CONF example

A sample SNMPD.CONF file is shipped as /usr/lpp/tcpip/samples/snmpd.conf.

See z/OS Communications Server: IP Configuration Reference for more information about syntax.

The sample OSNMP.CONF file used by the **snmp** command contains entries that match the sample SNMPD.CONF data set. See “Configure the z/OS UNIX snmp command” on page 1256 for additional information on configuring the **snmp** command.

Note: By default, the SNMP agent and the **snmp** command send packets such that a VIPA address will be used as the originating address in the packet, if SOURCEVIPA is configured. This is a change introduced in V2R10; previously, the SNMP agent and the **snmp** command set a socket option to cause the physical interface addresses to be used as the originating addresses on packets they sent. That meant the SNMPD.CONF file had to contain all of the possible physical interface addresses that might be used, rather than a smaller number of VIPA addresses. A customer can override this change in behavior, if wanted, by invoking the SNMP agent with the `-a` option or by using either the `-a` option or the NOSVIPA option in the **snmp** command's OSNMP.CONF configuration file.

SNMPD.BOOTS

The SNMP agent uses the SNMPD.BOOTS configuration file to support SNMPv3 security. This file contains agent information used to authenticate the SNMPv3 requests. The SNMPD.BOOTS keeps the agent identifier and the number of times the agent reboots. If no SNMPD.BOOTS file exists when the agent is started, the agent creates one. You may want to add comments to the beginning of this file. If a file does exist, the agent uses the values specified in the file for setting its engineID and engineBoots values. If the file exists but contains incorrect values for engineID or engineBoots, the agent issues a message and terminates.

Notes:

1. The recommended approach is to allow the SNMP agent to create the file.
2. If the SNMPD.BOOTS file is not provided, the SNMP agent creates the file. If multiple SNMPv3 agents are running on the same MVS image, use the environment variable to specify different SNMPD.BOOTS files for the different agents. For security reasons, ensure unique engineIDs are used for different SNMP agents.

Creating user keys

- Authentication

  Authentication is generally required for SNMPv3 requests to be processed (unless the security level requested is 'noAuth'). When authenticating a request, the SNMP agent verifies that the authentication key sent in an SNMPv3 request can be used to create a message digest that matches the message digest created from the authentication key defined for the user.

  The **snmp** command uses the authentication key found on an entry in the OSNMP.CONF configuration file. It needs to correlate with the authentication key specified on a USM_USER entry for that user in the agent's SNMPD.CONF configuration file.

  As an alternative to storing authentication keys in the client configuration file, the **snmp** command allows user passwords to be stored. If the **snmp** command is configured with a password, the code generates an authentication key (and privacy key if requested) for the user. These keys must, of course, produce the same authentication values as the keys configured for the USM_USER in the agent's SNMPD.CONF file or configured dynamically with SNMP SET commands. However, the use of passwords in the client configuration file is considered less secure than the use of keys in the configuration file.

  The authentication key is generated from two pieces of information:

  – The specified password.
  – The identification of the SNMP agent at which the key will be used. If the agent is an IBM agent and its engineID was generated using the vendor-specific engineID formula, the agent may be identified by IP address or host name. Otherwise, the engineID must be provided as the agent identification.

  A key that incorporates the identification of the agent at which it will be used is called a localized key. It can be used only at that agent. A key that does not incorporate the engineID of the agent at which it will be used is called nonlocalized.

  Keys stored in the **snmp** command's configuration file, OSNMP.CONF, are expected to be nonlocalized keys. Keys stored in the SNMP agent's configuration file, SNMPD.CONF, can be either localized or nonlocalized, though the use of localized keys is considered more secure.

- Encryption

  Keys used for encryption are generated using the same algorithms as those used for authentication. However, key lengths might differ. For example, an HMAC-SHA authentication key is 20 bytes long, but a localized encryption key used with HMAC-SHA is only 16 bytes long. The SNMP agent, z/OS UNIX **snmp** command, and the SNMP manager API use the first 16 bytes of the HMAC-SHA authentication key as the localized encryption key (also called the privacy key).

  z/OS Communications Server provides a facility called *pwtokey* that enables conversion of passwords into localized and nonlocalized authentication and privacy keys. The pwtokey procedure takes as input a password and an identifier of the agent and generates authentication and privacy keys. Because the procedure used by the pwtokey facility is the same algorithm used by the **snmp** command, the person configuring the SNMP agent can generate appropriate authentication and privacy keys to put in the SNMPD.CONF file for a user, given a particular password and the IP address at which the agent will run.

  Use the **pwtokey** command to convert passwords into authentication and privacy keys. See z/OS Communications Server: IP System Administrator's Commands.

Migrating community-based configuration to SNMPD.CONF format

If you want to continue to use community-based security but take advantage of some of the new SNMPv3 functions, or if you want to use the new SNMPv3 user-based security along with community-based security, you need to migrate your current configuration, defined in PW.SRC and SNMPTRAP.DEST, to SNMPD.CONF format. For information about the necessary steps for migrating PW.SRC and SNMPTRAP.DEST to SNMPD.CONF, see z/OS Communications Server: IP Configuration Reference.

Provide secure access to agent from subagents

An SNMP subagent can connect to the z/OS Communications Server SNMP agent by using the DPI API (the DPI API is documented in z/OS Communications Server: IP Programmer's Guide and Reference) and specifying either a z/OS UNIX or a TCP connection. You can control access to the agent from subagents for both types of connections.

Connecting to the agent through z/OS UNIX

For subagents that specify a z/OS UNIX connection to the agent, a z/OS UNIX path name is used for the connection. You can configure this path name by either specifying it on the `-s` agent initialization parameter or specifying it as the value of the dpiPathNameForUnixStream MIB object in OSNMPD.DATA. The default path name is /var/dpi_socket.

The SNMP agent creates this path name every time it initializes. For subagents to successfully connect to the agent using this path name, either the subagents must be defined with superuser authority or the read and write file access permission bits for the path name must be set as follows:

- If the user ID of a subagent is associated with the same security product group as the agent, read and write access must be set in the Group section of the file access permission bits.
- If the user ID of a subagent is not associated with the same security product group as the agent, read and write access must be set in the Other section of the file access permission bits.

You can configure the read and write access by specifying the -C agent initialization parameter.

For more detailed information about file access permission bits and handling security for your files, see z/OS UNIX System Services User's Guide. If you need to change the file access permission bits for the path name after the agent has initialized, you can use the z/OS UNIX **chmod** command. For more information about the **chmod** command, see z/OS UNIX System Services Command Reference.

Connecting to the agent through TCP

For subagents specifying a TCP connection, you can use the installation's SAF-compliant security product [such as the z/OS Security Server (RACF)] to control which of the SNMP subagents are permitted to connect to the SNMP agent. One security product resource name can be created per TCP/IP stack per MVS image. The security product resource name is specified in the following format:

```
EZB.SNMPAGENT.sysname.tcpprocname
```

where *sysname* is the name of the MVS system image and *tcpprocname* is the TCP/IP started procedure name.

The profile must be created under the SERVAUTH class. After creating the profiles, use the security product to define the user IDs of those subagents which should be permitted to connect via TCP to the SNMP Agent. Authorization failures are documented by security product failure messages and SNMP agent traces.

Note: If you use this authorization function, only SNMP subagents which are associated with the same TCP/IP stack as the SNMP agent will be permitted to connect to the agent. Local SNMP subagents that are associated with other TCP/IP stacks, or remote SNMP subagents, will not be permitted to connect. The following ICH0408I RACF error message is issued:

```
ICH408I JOB(OSNMPD  ) STEP(OSNMPD  )
  EZB.SNMPAGENT.sysname.tcpprocname CL(SERVAUTH)
  INSUFFICIENT ACCESS AUTHORITY
  FROM EZB.SNMPAGENT.** (G)
  ACCESS INTENT(READ   )  ACCESS ALLOWED(NONE   )
```

In the ICH0408I message, only the JOB and STEP names are displayed instead of a user ID. Because the subagent was not associated with the same TCP/IP stack as the SNMP agent, the agent could not obtain the user ID.

Also, any subagents which connected to the SNMP agent before the agent security product resource name was created will not have been authorized via the security product.

You can use the control statements in the sample JCL job provided in SEZAINST(EZARACF) to define this authorization. For example, if you wanted to permit any SNMP subagents associated with a user ID of USER2 to connect to the SNMP agent you could use the following definitions:

```
RDEFINE SERVAUTH EZB.SNMPAGENT.MVSA.TCP1 UACC(NONE)
PERMIT EZB.SNMPAGENT.MVSA.TCP1  ACCESS(READ) CLASS(SERVAUTH) ID(USER2)
```

Allowing subagents with duplicate identifiers to connect

When an SNMP subagent connects to the SNMP agent, an SNMP object identifier (OID) value is used as the subagent identifier. For subagents connecting using the DPI V2.0 API, the subagent supplies its OID during the connection process. All the Communications Server subagents use the DPI V2.0 API when connecting to the agent.

For subagents connecting using the DPI V1.1 API, the agent assigns a constant OID value to the subagent. Therefore, if more than one DPI V1.1 subagent connects to the agent, all the DPI V1.1 subagents are identified with the same OID value.

The SUBAGENT-MIB, supported by the SNMP agent, defines the MIB object, saAllowDuplicateIDs, that can be used to configure whether the agent should allow subagents with duplicate OID values to connect. By default, the SNMP agent sets the value of this object to 1, which allows subagents with duplicate OID values to connect. You can configure the value of this MIB object using the OSNMPD.DATA configuration file. For more information about this file, see “Provide MIB object configuration information” on page 1253. If you do not want the agent to allow subagents with duplicate OID values to connect, set the value of this MIB object to 2.

Provide MIB object configuration information

An installation can set values for selected MIB objects by providing OSNMPD.DATA information. A sample of OSNMP.DATA is installed as file /usr/lpp/tcpip/samples/osnmpd.data. See z/OS Communications Server: IP Configuration Reference for syntax information. If no OSNMPD.DATA file is found, the defaults for these MIB objects are as follows:

Object
Default

dpiPathNameForUnixStream
The default is /var/dpi_socket. This is the z/OS UNIX path name that is used in accepting requests from subagents that communicate with the agent over z/OS UNIX connections.

The SNMP agent creates this path name every time it initializes. For subagents to successfully connect to the agent using this path name, either the subagents must be defined with superuser authority or the read and write file access permission bits for the path name must be set as follows:

- If the user ID of a subagent is associated with the same security product group as the agent, read and write access must be set in the Group section of the file access permission bits.
- If the user ID of a subagent is not associated with the same security product group as the agent, read and write access must be set in the Other section of the file access permission bits.

You can configure the read and write access by specifying the -C agent initialization parameter.

For more detailed information about file access permission bits and handling security for your files, see z/OS UNIX System Services User's Guide. If you need to change the file access permission bits for this path name after the agent has initialized, you can use the z/OS UNIX **chmod** command. For more information about the **chmod** command, see z/OS UNIX System Services Command Reference.

sysDescr
If the environment variable HOSTNAME exists, its value is used. Otherwise, the default value identifies the z/OS system under which the agent is running. The maximum length of this object is 255 octets.

sysContact
"SNMPBASE-Unspecified". The maximum length of this object is 255 octets.

sysLocation
"SNMPBASE-Unspecified". The maximum length of this object is 255 octets.

sysName
"SNMPBASE-Unspecified". The maximum length of this object is 255 octets.

sysObjectId
1.3.6.1.4.1.2.3.13

Note: sysObjectID is defined as the vendor's authoritative identification of the network management subsystem contained in the entity. That is, it is intended to uniquely identify the SNMP agent. Changing this value is not recommended and will be disabled in a subsequent release.

sysServices
A single octet with a default of 0. See the RFC 1907 description for this object.

snmpEnableAuthenTraps
Default value is 2, which means traps are disabled.

saDefaultTimeout
5 seconds.

saMaxTimeOut
600 seconds.

saAllowDuplicateIDs
Default is 1, which means multiple instances of a subagent (that is, where the subagent identifier for all the subagents is the same) are allowed to connect to the SNMP agent. To prevent multiple instances of a subagent from connecting to the SNMP agent, set the value to 2.

Note: Because a subagent identifier cannot be specified for subagents connecting using the DPI V1.1 API, the SNMP agent assigns the same constant identifier for all DPI V1.1 subagents. Therefore, this object must be set to 1 to allow multiple DPI V1.1 subagents to run concurrently. For more information about subagent identifiers, see "Allowing subagents with duplicate identifiers to connect" on page 1253.

For information about where these MIB objects are defined, see z/OS Communications Server: IP User's Guide and Commands.

If creating a data set, you can specify a sequential data set with the following attributes: RECFM=FB, LRECL=80, and BLKSZ=3120. Other data set attributes might also work, depending on your installation parameters.

Common INET considerations

If you have configured a common INET (CINET) environment where multiple TCP/IP stacks are active, and you want to run multiple SNMP agents, each associated with a specific stack, review the following additional configuration steps:

- You probably need to create a separate set of SNMP agent configuration files for each SNMP agent that you are going to start.
- If you want to start SNMP subagents that connect to the agent using a z/OS UNIX path name, then you need to perform the following tasks:
  - Define a unique path name for each stack.
  - Specify this path name on the SNMP agent `-s` start parameter.

    See OSNMPD parameters in z/OS Communications Server: IP Configuration Reference for more information about this parameter.

  All of the z/OS Communications Server SNMP subagents connect to the agent using a z/OS UNIX path name.

Start the SNMP agent

The SNMP agent (OSNMPD) runs in a separate address space that executes load module EZASNMPD. OSNMPD can be started with or without parameters. When starting OSNMPD from MVS, add the parameters to the PARMS= keyword on the EXEC statement of the OSNMPD cataloged procedure. When starting OSNMPD from z/OS UNIX, specify the parameters on the **osnmpd** command. See OSNMPD parameters in z/OS Communications Server: IP Configuration Reference for the command syntax.

For information about starting the SNMP agent in a common INET (CINET) environment, see "Common INET considerations" on page 1254.

If the SNMP agent encounters any errors processing its configuration files, error messages are written to the syslog daemon, not to the console.

Sample JCL procedure for starting OSNMPD from MVS

Update cataloged procedure OSNMPD by copying the sample in SEZAINST(OSNMPDPR) to your system or recognized PROCLIB. Change the data set names as required to suit your local configuration. The OSNMPD address space requires access to the z/OS XL C/C++ run-time library data sets during execution.

You can pass parameters to the agent on the PARMS= keyword on the EXEC statement of the OSNMPD cataloged procedure. See z/OS Communications Server: IP Configuration Reference for the command syntax and parameter information. You can specify any agent parameters that you want, as shown in the following example:

```
//OSNMPD EXEC PGM=EZASNMPD,REGION=4096K,TIME=NOLIMIT,
// PARM='-c abc -d 255 -p 761'
```

In this example, the agent will use port 761 to accept requests, community name abc will be added to the list of community names supported by the agent, and all agent traces will be activated. For more information on tracing, see the z/OS Communications Server: IP Diagnosis Guide.

Starting OSNMPD from z/OS UNIX

To run the SNMP agent in background, you must add an ampersand (&) to the command and the issuer of the command must be in z/OS UNIX superuser mode. For a detailed explanation of the parameters, see z/OS Communications Server: IP Configuration Reference.

Any agent parameters that you want to specify can be added as shown in the following example:

```
osnmpd -c abc -d 255 -p 761
```

In this example, the agent will use port 761 to accept requests, community name 'abc' will be added to the list of community names supported by the agent, and all agent traces will be activated. For more information on tracing, see z/OS Communications Server: IP Diagnosis Guide.

Step 2: Configure the SNMP commands

The two SNMP client applications provided with z/OS Communications Server are:

- **snmp** command in the z/OS shell
- SNMP command from the NetView environment

The SNMP command in the NetView environment requires the use of the NetView product. It supports SNMP version 1. The **snmp** command in the z/OS shell supports SNMP versions 1, 2, and 3. Depending on your requirements, you might decide to configure either or both of these clients, or to use an SNMP client on another platform.

Configure the z/OS UNIX snmp command

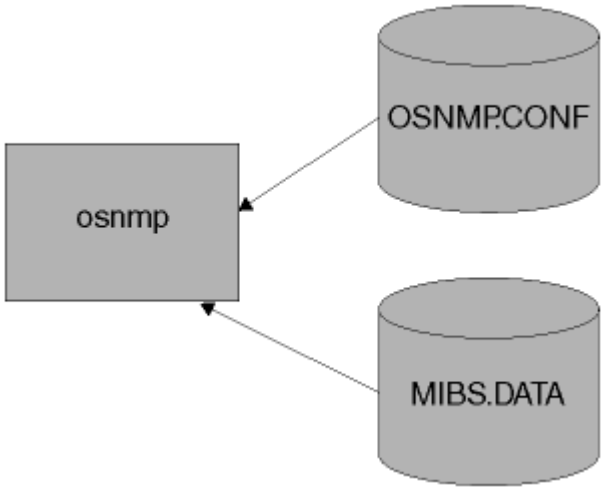


*Figure 153. Configuration files for **snmp***

The z/OS UNIX **snmp** command is used to send SNMP requests to SNMP agents on local or remote hosts, or to receive SNMP traps or notifications. You can also use the synonym, **osnmp**, as the name of this command. The requests can be SNMPv1, SNMPv2c, or SNMPv3. For SNMPv2c and SNMPv3 requests, the OSNMP.CONF configuration file is required. The *winSNMPname* specified on an OSNMP.CONF statement can be used as the value of the -h parameter on the **snmp** command. For a detailed explanation of the parameters you can specify on the **snmp** command, see z/OS Communications Server: IP System Administrator's Commands.

To configure the **snmp** command, perform the following tasks:

- The **snmp** command needs to be able to resolve the name of the host, on which the command is executing, to the host's IP address. You can provide this information either by configuring a domain name server or by configuring local host files. For information about the search order used to locate local host files, see the local host tables entry in Table 36 on page 772. As part of its support for the SNMPv3 protocol, the command also uses this IP address when creating its SNMPv3 engineID value.
- Provide **snmp** configuration information
- Provide user MIB object information

Provide snmp configuration information

The OSNMP.CONF file is used to define target agents and, for SNMPv3, the security parameters to be used in sending requests to them.

The contents of the file, regardless of location, are the same. Only the first file found is used. A sample of this file is installed as file /usr/lpp/tcpip/samples/snmpv2.conf. This sample should be copied and modified for your installation. See z/OS Communications Server: IP Configuration Reference for more information.

Examples

- Example 1:

  The following entry defines an SNMPv2c node for **snmp**:

  ```
  mvs1     9.67.113.79 snmpv2c
  ```

  where *mvs1* is the name used with the -h parameter on the **snmp** command and *9.67.113.79* is the IP address of the SNMPv2c agent.

- Example 2:

  The following entry defines an SNMPv3 node:

```
v3mak   127.0.0.1 snmpv3 u1 - - AuthNoPriv  HMAC-MD5 7a3e34265e0e029f27d8b4235ecfa987 - -
```

where *v3mak* is the name used on the -h parameter of the **snmp** command. An SNMP request sent using this entry uses USM user name *u1* using HMAC-MD5 authentication but no encryption.

- Example 3:

  The following entry defines an SNMPv3 node. The needed authentication and privacy keys will be generated from the password *u6password*.

  ```
  v3sap   127.0.0.1 snmpv3 u6 u6password - AuthNoPriv   HMAC-SHA -   - -
  ```

  The USM user is *u6*. The authentication protocol is HMAC-SHA, and no encryption is used.

- Example 4:

  The following entry defines an SNMPv3 node:

  ```
  v3mpk_ipv6 ::1 snmpv3 u1 - - AuthPriv HMAC-MD5 7a3e34265e0e029f27d8b4235ecfa987 DES eac02a0d9fe90eca7911fdcaba20deae
  ```

  where *v3mpk_ipv6* is the name used on the -h parameter of the **snmp** command. An SNMP request sent using this entry uses USM user name *u1* using HMAC-MD5 authentication and CBC 56-bit DES encryption.

Provide MIB object information in MIBS.DATA

Like other SNMP managers, when you enter the z/OS UNIX **snmp** command you can specify the object identifier (OID) of the MIB object whose value you want to retrieve. If the MIB object is supported by one of Communication Server's SNMP functions, you can also use the MIB object textual name on the z/OS UNIX **snmp** command, instead of the object identifier. All the MIB objects supported by Communications Server functions are listed in the MIB objects appendix in z/OS Communications Server: IP System Administrator's Commands.

For example, to retrieve the SNMP agent sysUpTime MIB object, you can enter either of the following commands:

- ```
  snmp -v get sysUpTime.0
  sysUpTime.0 = 289700
  ```

- ```
  snmp -v get 1.3.6.1.2.1.1.3.0
  sysUpTime.0 = 292700
  ```

The 1.3.6.1.2.1.1.3 value is the OID for the sysUpTime MIB object.

If you have user-defined MIB objects or MIB objects from other products, and you want to use the textual names for these MIB objects on the z/OS UNIX **snmp** command, you can define these objects to the **snmp** command using the MIBS.DATA file. A sample of the MIBS.DATA file is installed as file /usr/lpp/tcpip/samples/mibs.data. Copy this sample and modify it for your installation. For the search order used to locate the MIBS.DATA file, see "Understanding search orders of configuration information" on page 13.

If other products provide files using MIBS.DATA syntax, append the statements from their files to your MIBS.DATA file so that these statements are accessible to the **snmp** command. For example, the IBM OSA-Express Direct subagent provides a file in MIBS.DATA syntax so that the textual names of its OSA-Express management data can be used with the z/OS UNIX **snmp** command. For information about how to obtain this file, see Open Systems Adapter-Express Customer's Guide and Reference.

MIBS.DATA statement syntax

The format of a statement in the MIBS.DATA file is:

```
character_object_name object_identifier object_type
```

See z/OS Communications Server: IP Configuration Reference for more information about syntax.

Configure the NetView SNMP command

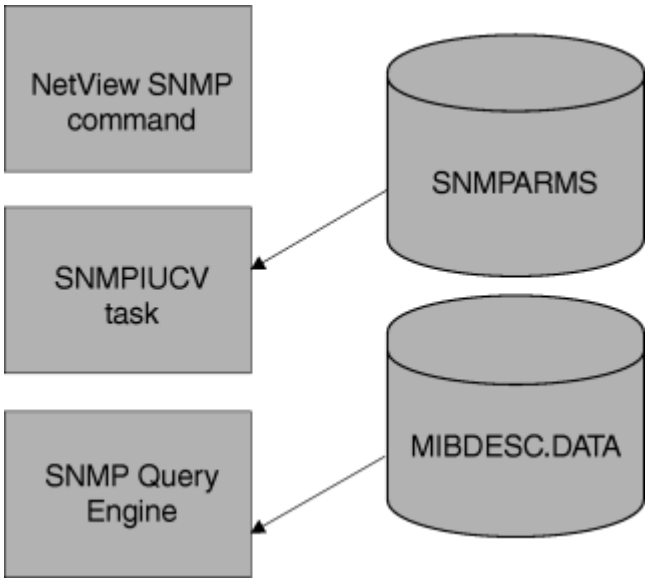


Figure 154. Configuration files for NetView SNMP

The SNMP command in the NetView environment can be used to send SNMP version 1 requests to SNMP agents on either local or remote hosts. The SNMP command requires the command processor itself, the SNMPIUCV task for inter-address space communication, and the SNMP query engine, which creates the packets sent to the SNMP agent. The NetView SNMP command and the SNMP query engine support only community-based security.

Configure the SNMP query engine

Update the SNMPQE cataloged procedure by copying the sample in SEZAINST(SNMPPROC) to your system or recognized PROCLIB. Specify SNMP parameters and change the data set names as required to suit your local configuration. The SNMPQE address space requires access to the z/OS XL C/C++ run-time library data sets during execution.

The SNMP query engine (SQESERV) needs access to the *hlq*.MIBDESC.DATA data set for the MIB variable descriptions. You can find a sample of this data set in SEZAINST(MIBDESC).

MIBDESC.DATA data set

The MIBDESC.DATA data set defines the short names for MIB variables. Short names are the character representation for the ASN.1 variable names. For example, sysUpTime is the short name for 1.3.6.1.2.1.1.3.0 (the MIB variable that stores the time since the SNMP agent was last restarted). Short names are generally shown as a combination of upper and lowercase characters, though SNMP on z/OS Communications Server ignores these case distinctions. Variable names must always be in ASN.1 language when they are sent to an SNMP agent. You can always use ASN.1 language to specify the variable names in an enterprise-specific tree (assuming that the agent supports them). You can use these short names to specify the MIB variables.

When you issue an SNMP GET, GETNEXT, or SET command, and specify the variable name in ASN.1 notation, the SNMP Query Engine uses that name and sends it in the SNMP packet to the agent. When you issue an SNMP GET, GETNEXT, or SET command, and specify the short name for the variable (for example, sysDescr), the SNMP Query Engine looks for that name in the MIBDESC.DATA data set and uses the ASN.1 name specified in the data set when it sends the SNMP packet to the agent.

The SNMPQE address space must be able to access the MIBDESC.DATA data set.

You can change the short names in the MIBDESC.DATA data set to the equivalent in your national language. You can also leave the current names and add the equivalent names in your national language. However, the SNMP MIBVNAME function returns only the first entry found in the data set that satisfies the

search. In addition, all enterprise-specific variables used by hosts in your network should be added to this data set.

Entries in the data set do not need to be in a specific sequence. Each name starts on a new line. The following entry shows the line format.

```
short_name asn.1_name type time_to_live
```

Each variable on the line is separated by either one or more spaces or tabs. An asterisk (*) in column 1 indicates that the line is a comment line.

The following sample is a MIBDESC.DATA line with a sysDescr variable translated in Dutch and a few enterprise variables added (in this example, company ABC received 1.3.6.1.4.1.42 as the ASN.1 number for their enterprise):

```
*-------------------------------------------------------------*
* MIB Variable name | ASN.1 notation        | Type     | TTL  *
*-------------------------------------------------------------*
* Following is Dutch name for sysDescr
systeemBeschrijving  1.3.6.1.2.1.1.1.         display    900
sysDescr             1.3.6.1.2.1.1.1.         display    900
  ...
  other entries
  ...
* Following are Enterprise-Specific variables for company ABC
ABCInfoPhone         1.3.6.1.4.1.42.1.1       display    900
ABCInfoAddress       1.3.6.1.4.1.42.1.2       display    900
```

The TTL field contains the number of seconds that a variable lives in the Query Engine's internal cache. If there are multiple requests for the same variable within the TTL period, the variable value is obtained from the cache, and unnecessary network traffic is avoided.

You can define multiple short names or text names for the same variable, as shown with the Dutch translation of the sysDescr variable. In this case, the SNMP Query Engine returns the first value in the table on an SNMP MIBVNAME request. In the previous example, the SNMP Query Engine would return `systeemBeschrijving` and not `sysDescr`. The name returned is in mixed case.

When the SNMP Query Engine receives a short name or text name in a GET, GETNEXT, or SET request, it compares the name against the entries in the MIBDESC.DATA data set. This comparison is not case-sensitive. For example, a request for SYSDESCR, `SysDescr`, or `sysDescr` matches the `sysDescr` entry with an ASN.1 notation of `1.3.6.1.2.1.1.1.`.

When the SNMP Query Engine receives an SNMP response, it looks up the variable in the MIBDESC.DATA table Type field for information about translating the value into displayable characters. The information contained in the Type field is case-sensitive and must be specified in lowercase.

Note: If you are using SNMP to receive response or trap PDUs which contain enterprise-specific variables, the variables must be added to the MIBDESC.DATA data set.

Specifying the SNMPQE parameters

The SQESERV module can be configured to start without parameters or you can add any of the following parameters to PARMS=' in the PROC statement of the SNMPQE cataloged procedure. For example,

```
//SNMPQE   PROC MODULE=SNMPQE,PARMS='-h MVSA'
```

See z/OS Communications Server: IP Configuration Reference for the command syntax.

See z/OS Communications Server: IP Diagnosis Guide for more information on tracing.

Setting up authorization for SNMPQE

To create RAW sockets necessary for SNMP PING requests, the user ID associated with the SNMPQE started task must have superuser authority (z/OS UNIX UID of 0), or must be permitted to become a superuser by having READ access to the BPX.SUPERUSER resource in the FACILITY class.

Configure NetView as an SNMP monitor

To configure the NetView interface as an SNMP monitor, perform each of the following tasks:

- Configure for SNMPIUCV
- Configure for the SNMP command processor
- Configure for the SNMP messages
- Update the SNMP initialization parameters

Configure for SNMPIUCV

SNMPIUCV is the NetView optional task that handles IUCV communication between the NetView program and the SNMP query engine. SNMPIUCV is in the SEZADSIL data set.

Add the following TASK statement for SNMPIUCV to the DSIDMN member of the data set specified by the DSIPARM DD statement in the NetView start procedure.

```
TASK  MOD=SNMPIUCV,TSKID=SNMPIUCV,PRI=5,INIT=Y
```

This statement causes SNMPIUCV to start automatically when the NetView program is started.

If you specify `INIT=N` instead of `INIT=Y` in the TASK statement for SNMPIUCV, a NetView operator can start the SNMPIUCV task by entering the following command:

```
START TASK=SNMPIUCV
```

The SNMPIUCV task tries to connect through IUCV to the SNMP query engine. If this fails, it tries the connect again as specified by the SNMPQERT keyword in the SNMPARMS member of the SEZADSIP data set. The default is every 60 seconds.

Configure for the SNMP command processor

SNMP is the command processor that allows NetView operators and CLISTs to issue SNMP commands. SNMP is in the SEZADSIL data set. This data set should be concatenated to the STEPLIB DD statement in the NetView start procedure.

Add the following statement to the DSICMD member of the data set specified by the DSIPARM DD statement in the NetView start procedure.

```
SNMP  CMDMDL  MOD=SNMP,ECHO=Y,TYPE=R,RES=Y
```

After the SNMPIUCV task is started, you can issue the SNMP command. The SNMP command passes a request to the SNMPIUCV task to forward to SNMPQE. The return code represents a request number that is associated with the request. The responses are returned asynchronously and contain this request number. The operator or CLIST must use the request number to correlate the response to the request.

Configure for the SNMP messages

The NetView SNMP messages are in the SEZADSIM data set as DSISNM*nn*, where *nn* is the number of the member. The valid message members are DSISNM00 through DSISNM05, DSISNM10, DSISNM12, and DSISNM99. The data set containing these members should be added to the DSIMSG DD statement in the NetView start procedure.

Update the SNMP initialization parameters

SNMPIUCV reads the SNMPARMS member in the SEZADSIP data set at startup. This data set contains the initialization parameters for SNMP. The data set containing SNMPARMS should be added to the DSIPARM DD statement in the NetView startup procedure. See z/OS Communications Server: IP Configuration Reference for detailed information for the SNMP parameter data set (SNMPARMS).

Step 3: Configure the SNMP subagents

There are several SNMP subagents shipped with z/OS Communications Server:

- The TCP/IP subagent reports information about the TCP/IP stack. For details on configuring this subagent, see "TCP/IP subagent configuration" on page 1261.
- The OMPROUTE subagent reports information specific to OSPF. The ROUTESA_CONFIG statement is used in the OMPROUTE configuration file to configure the OMPROUTE subagent. For details on ROUTESA_CONFIG, see z/OS Communications Server: IP Configuration Reference.
- The Network SLAPM2 subagent reports information about defined policies and performance statistics related to traffic using those policies. For configuration information about the Network SLAPM2 subagent, see Chapter 15, "Quality of service," on page 831.
- The TN3270E Telnet subagent reports information about TN3270E Telnet server connections and monitoring values. For information on the TNSACONFIG statement and the parameters needed to start the TN3270E Telnet subagent, see z/OS Communications Server: IP Configuration Reference.
- The OSA-Express Direct SNMP subagent is also shipped with z/OS CS but is supported by the System z OSA Support Group. The OSA-Express Direct SNMP subagent and the OSA MIB provided by the System z OSA Support Group can be used with Communications Server SNMP support to provide SNMP management data for some OSA adapters. For details regarding the OSA-Express Direct SNMP subagent and OSA MIB, see Open Systems Adapter-Express Customer's Guide and Reference.

These subagents use the stack's IPv4 primary interface IP address when connecting to the SNMP agent. The IPv4 primary interface IP address is either the first IPv4 IP address specified in the HOME profile statement, or the IP address of the IPv4 interface specified on a PRIMARYINTERFACE profile statement. If the subagents cannot obtain the IPv4 primary interface IP address, they will use the IPv4 loopback IP address, 127.0.0.1, to connect to the agent. For information on using the IPv4 primary interface IP address and a community name to permit the subagents to connect to the SNMP agent, see "Provide community name information" on page 1247 or "Provide community-based and user-based security and notification destination information" on page 1249.

TCP/IP subagent configuration

There are two statements in the profile data set used to configure the TCP/IP subagent, the SACONFIG and ITRACE statements.

- SACONFIG

  Use the SACONFIG statement to configure the subagent. The SACONFIG parameters determine whether or not the subagent is automatically started at TCP/IP initialization, what port number to use to contact the agent, and other configuration values. For a detailed explanation of this statement, see z/OS Communications Server: IP Configuration Reference.

- ITRACE

  Use the ITRACE statement to determine what trace information, if any, should be recorded by the subagent. For a detailed explanation of this statement, see z/OS Communications Server: IP Configuration Reference.

Step 4: Configure the Open Systems Adapter support

The TCP/IP subagent can retrieve SNMP management data from the Open Systems Adapter Support Facility (OSA/SF) for several OSA adapters.

The OSA product also provides an SNMP subagent, the OSA-Express Direct subagent, that supports management data for OSA-Express adapters. You can use the OSA-Express Direct subagent with Communications Server SNMP support to retrieve the management data. Use the OSA-Express Direct subagent for OSA management data, rather than the TCP/IP subagent, for the following reasons:

- The OSA-Express Direct subagent communicates directly with the OSA-Express adapters and does not require the OSA/SF and IOASNMP applications.

- The OSA-Express Direct subagent supports all OSA-Express adapters. The TCP/IP subagent does not support OSA-Express3 or later adapters.

For more information about the management data provided by the OSA-Express Direct subagent, see Open Systems Adapter-Express Customer's Guide and Reference.

The TCP/IP subagent supports management data for the following OSA-Express2 adapters:

- Ethernet management data is supported for Gigabit Ethernet or QDIO Fast Ethernet adapters.
- Tables of OSA-Express management data are supported for Gigabit Ethernet or Fast Ethernet adapters.

Restriction: The TCP/IP subagent does not support OSA-Express3 or later adapters.

The TCP/IP subagent retrieves the data using the OSA/SF components IOAOSASF (OSA/SF application) and IOASNMP (OSA/SF socket application). For more information on these components, see "OSA/SF prerequisites" on page 1263. Specifying the SACONFIG profile statement, with the OSAENABLED and OSASF parameters, in the TCP/IP profile data set causes the TCP/IP subagent to try to connect to the OSA/SF socket application, IOASNMP, using the TCP protocol and the port number specified on the OSASF parameter. If the subagent connects successfully, the following message is issued:

```
EZZ3218I SNMP SUBAGENT: CONNECTED TO OSA/SF
```

Because of timing considerations, the TCP/IP subagent might not be able to connect to IOASNMP at initalization. If this occurs, the subagent will attempt to connect when the first request for OSA MIB data is received. Therefore, the EZZ3218I message might not always be issued during subagent initialization.

When retrieving management data from the OSA adapters, the TCP/IP subagent sends a request to IOASNMP for the data, passing the adapter's portname as an identifier. The portname is obtained from the DEVICE and LINK profile statements used to define the adapter to the TCP/IP stack. The IOASNMP socket application uses APPC to pass the request to the OSA/SF application. The OSA/SF application then retrieves the data from the adapter and returns it to IOASNMP. IOASNMP uses its TCP connection to the TCP/IP subagent to return the data to the subagent. If this configuration is active and either the IOASNMP application or the TCP/IP subagent terminates, the subagent will issue the following message:

```
EZZ3219I SNMP SUBAGENT: DISCONNECTED  FROM OSA/SF
```

To obtain the management data, the adapters must be defined to the TCP/IP stack where the subagent is active, through DEVICE and LINK statements in the TCP/IP profile.

- To retrieve Ethernet management data, the OSA-Express adapter must be defined as an MPCIPA Ethernet link.
- To retrieve OSA-Express management data, the adapters must be defined as the following TCP/IP device/link types:
  - MPCIPA link for Gigabit Ethernet
  - MPCIPA or LCS Ethernet link for Fast Ethernet

If the port name is manually configured at the adapter, then management data can be retrieved from the adapter even if it is not active and not in use by any TCP/IP stack or by VTAM. If the port name is dynamically configured (e.g. MPCIPA links), then the adapter has to be active to some TCP/IP stack to retrieve the management data.

Current support consists of:

- OSA-Express Channel and Performance tables from the IBM MVS Enterprise-Specific MIB for OSA-Express Gigabit Ethernet and Fast Ethernet adapters. The performance MIB object values in the ibmMvsOsaExpChannelTable, and all of the MIB object values in the ibmMvsOsaExpPerfTable, are only available from OSA-Express adapters on which a microcode level of 1.31 or later is installed, and that are active on a System z processor.
- OSA-Express Ethernet Port table from the IBM MVS Enterprise-Specific MIB for OSA-Express Gigabit Ethernet and Fast Ethernet adapters.

- OSA-Express Ethernet SNA table from the IBM MVS Enterprise-Specific MIB for OSA-Express Fast Ethernet adapters.
- dot3StatsTable from EtherLike-MIB in RFC2665 for OSA-Express Gigabit Ethernet and QDIO Fast Ethernet adapters.

OSA/SF prerequisites

The TCP/IP subagent provided by z/OS Communications Server will connect to OSA/SF to obtain management data. For a subagent to establish a connection to OSA/SF, two OSA/SF components must be started:

- IOAOSASF

  IOAOSASF is a sample JCL procedure that can be used to start the main OSA/SF address space. The sample has a job name of OSASF1.

- IOASNMP

  IOASNMP is a sample JCL procedure that starts the OSA/SF-provided z/OS UNIX socket application that interconnects the TCP/IP subagent with OSASF1.

These sample procedures and all entities that they call are provided with OSA/SF. For a detailed explanation of how to set up OSA/SF on your MVS system, see Open Systems Adapter-Express Customer's Guide and Reference. The primary purpose of OSA/SF is to manage OSA Adapters. It has been extended to support OSA management via SNMP. An instance of IOAOSASF, IOASNMP, the TCP/IP stack, the TCP/IP subagent, and the SNMP agent must be started on every MVS image where OSA management support is needed.

The recommended startup order is:

1. Start IOAOSASF and wait until it completely initializes. IOAOSASF must be started before IOASNMP.
2. Include OSNMPD (the CS SNMP agent) and IOASNMP on the AUTOLOG profile statement for your TCP/IP stack. This ensures that they will be started by autolog processing when TCP/IP is started. For additional profile statement requirements, see “Required TCP/IP profile statements” on page 1263. Start the TCP/IP stack.

On an MVS image only a single instance of either IOAOSASF or IOASNMP can (or should) be started. An attempt to start multiple copies of IOAOSASF will be rejected. Starting multiple copies of IOASNMP will yield unpredictable results.

Ensure that OSA/SF is at Version 2 Release 1 level or higher with the OSA/SF APAR OW45237 applied.

Required TCP/IP profile statements

For a detailed description of the statements mentioned here, see z/OS Communications Server: IP Configuration Reference. The following TCP/IP profile statements must be updated for OSA management support:

- SACONFIG

  On the SACONFIG statement, OSA Management support must be enabled by specifying OSAENABLED. Omission of OSAENABLED when TCP/IP is started will result in no OSA management support. The SACONFIG statement controls the operation of the subagent that runs in a TCP/IP address space as a separate task.

  The OSASF parameter specifies which port IOASNMP should use to Listen for connections from subagents to OSA/SF. For an explanation of the usage of this parameter when starting multiple TCP/IP instances, see “Subagent connection to OSA/SF when there are multiple TCP/IP instances” on page 1264. It is recommended that the OSASF port be reserved by also specifying it on a PORT statement.

  For example:

```
SACONFIG OSAENABLED OSASF 721
```

- PORT

  Prereserve the port to be used in communication with OSA/SF.

  For example:

```
PORT
   721  TCP IOASNMP ; OSA/SF TCP/IP Communications
```

  In this example, because IOASNMP runs as a z/OS UNIX application, the port could have been reserved for z/OS UNIX. Review the /etc/services file to ensure that there are no port conflicts.

- DEVICE and LINK

  Provide DEVICE and LINK statements for any OSA-Express Ethernet adapter for which you want SNMP Ethernet management data. For example:

```
DEVICE portname MPCIPA NONROUTER
LINK linkName IPAQENET portname
```

Subagent connection to OSA/SF when there are multiple TCP/IP instances

When multiple z/OS Communications Server instances are active in the same MVS image, only one of the TCP/IP instances is connected to IOASNMP. For a TCP/IP instance to connect to IOASNMP, the OSASF parameter must be specified on the SACONFIG profile statement.

IOASNMP connects to a TCP/IP instance and acts as a server, receiving connections from those TCP/IP subagents where OSAENABLED was specified on the SACONFIG Profile statement. The result is that all these subagents connect through the same TCP/IP to IOASNMP in order to retrieve OSA information from OSA/SF. For a depiction of this process, see Figure 155 on page 1264.

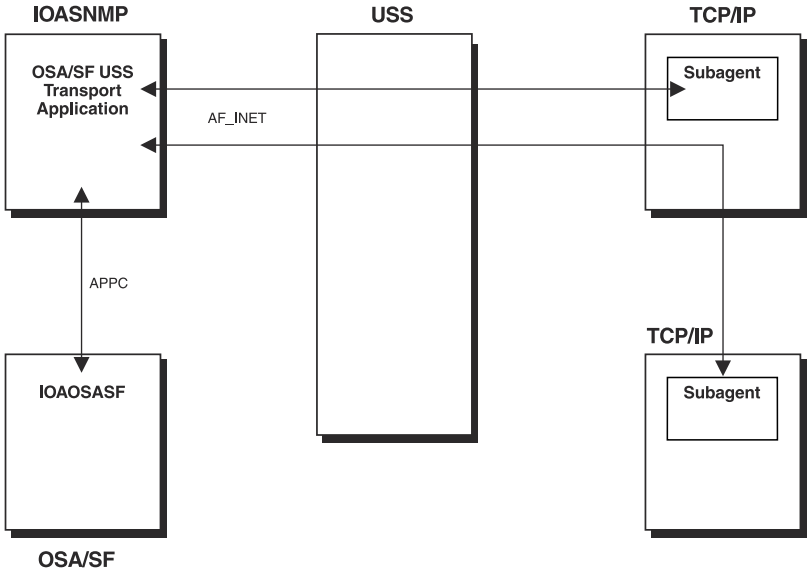


Figure 155. Subagent connection to OSA/SF

If IOASNMP loses its connection to TCP/IP it terminates and needs to be restarted.

If the currently connected TCP/IP terminates, IOASNMP will attempt to connect to another TCP/IP instance for which the OSASF parameter was specified on the SACONFIG Profile statement, using the port number specified for the OSASF parameter. The subagents will also attempt to reconnect to OSA/SF via IOASNMP using this same OSASF port number. For this reason it is recommended that the same OSASF port number be used on the SACONFIG statement of every TCP/IP instance where the OSASF parameter is specified.

Whenever a socket error occurs on the OSA/SF socket, the connected subagents will issue the following message:

```
EZZ3219I SNMP SUBAGENT: DISCONNECTED FROM OSA/SF
```

When the subagent connection is reestablished, the following message is issued:

```
EZZ3218I SNMP SUBAGENT: CONNECTED TO OSA/SF
```

Step 5: Configure the trap forwarder daemon

Most SNMPv1 agents forward traps to port 162. If a manager needs to listen for these traps, it has to bind to port 162 and listen for it. When a manager has already issued a bind it is impossible for another manager to listen for the same traps. The Trap Forwarder daemon eliminates this problem by listening for traps on port 162 and forwarding them to all the configured managers.

You can configure the Trap Forwarder daemon to receive a trap on a specified port and forward it to multiple other ports on the same host and on different hosts. This will allow multiple SNMP managers on z/OS to be able to receive all the traps sent to one port.

To configure the Trap Forwarder daemon, perform the following tasks:

1. Provide PROFILE.TCPIP statements.
2. Provide Trap Forwarder configuration.
3. Start the Trap Forwarder daemon (TRAPFWD).

Provide PROFILE.TCPIP statements

Add or update the following PORT configuration statements in *hlq*.PROFILE.TCPIP.

The default port used by the trap forwarder daemon to receive trap datagrams is UDP port 162. If you want to ensure that no other application uses this port, you must specify the following PORT statement:

```
PORT
  162    UDP    TRAPFWD           ;  Trap Forwarder daemon
```

If the daemon will be started from the z/OS shell, reserve the port for z/OS UNIX by changing OMVS instead of TRAPFWD. Note that by doing so, the **snmp** command could make use of the port if it is started (with the trap option) before TRAPFWD is started.

Provide trap forwarder configuration information

The TRAPFWD.CONF file defines the configuration parameters for trapfwd daemon.

A sample of the TRAPFWD.CONF is shown below:

```
#
# A sample configuration file for trapfwd
#
# Syntax  : ip_address/host_name            port_number           option
#
#
9.67.113.79               1064           ADD_RECVFROM_INFO
myHost                    1066
myHost                    1067           ADD_RECVFROM_INFO
myHost                    1099
9.67.35.37                1064           ADD_RECVFROM_INFO
-                         1065
-                         1068           ADD_RECVFROM_INFO
12ab::2                   1069
```

For more information about the statement syntax, see z/OS Communications Server: IP Configuration Reference.

Starting and stopping the trap forwarder daemon

The Trap Forwarder daemon can be started from the z/OS UNIX shell or from the MVS console.

Starting the trap forwarder daemon from z/OS UNIX

This example starts the Trap Forwarder daemon on the standard port (port 162).

```
# trapfwd
```

This example starts the Trap Forwarder daemon on a particular port (port 5062).

```
# trapfwd -p 5062
```

Starting the trap forwarder daemon from an MVS console

Update cataloged procedure TRAPFWD by copying the sample in SEZAINST(EZASNTRA) to your system. See z/OS Communications Server: IP Configuration Reference for more information about this cataloged procedure.

Stopping the trap forwarder daemon

From MVS:

```
P TRAPFWD
```

If TRAPFWD was started from a cataloged procedure, *procname* is the member name of that procedure. If TRAPFWD was started from the z/OS shell, *procname* is based on the *userid*. If the *userid* is 8 characters long, the *procname* is the *userid*. If the *userid* is less than 8 characters long, the *procname* is useridX, where X is the sequence number set by the system. To determine the sequence number issue **d omvs u=userid** from the console. This will show the programs running under the user ID.

From UNIX:

```
kill < pid >
```

Issue the **kill** command to the process ID (PID) associated with TRAPFWD. To find the PID issue the **ps -ef** command from the z/OS shell.

Tracing

The MODIFY command can be used to display the existing tracing level and also to change the tracing level.

The following example sets the trace level of the Trap Forwarder Daemon to 1.

```
MODIFY TRAPFWD,TRACE,LEVEL=1
```

The following example displays the level of tracing set for the Trap Forwarder Daemon.

```
MODIFY TRAPFWD,TRACE,QUERY
```

See z/OS Communications Server: IP Configuration Reference for more information about syntax.

Dynamically refreshing configuration

The MODIFY command can be used to dynamically refresh the configuration information. When this is done, the old configuration information is discarded completely. The configuration file is read again and the daemon is initialized.

The following example refreshes the configuration by reading the configuration file.

```
MODIFY TRAPFWD,REFRESH
```

See z/OS Communications Server: IP Configuration Reference for more information about syntax.

Chapter 24. Remote print server

Read “Understanding search orders of configuration information” on page 13. It covers important information about data set naming and search sequences.

The remote print server supports the line print daemon (LPD) and allows you to print on JES controlled printers from any host in your IP network that implements the line print client functions. These client functions are invoked with the LPR command. LPR is available as a TSO command, and the LPD server is implemented as a started z/OS task.

See z/OS Communications Server: IP System Administrator's Commands for information on starting and stopping the TCP/IP print server (LPD). When LPD is stopped by the MVS operator with the *P procname* command, LPD will terminate any TCP/IP connections currently transferring data. Before ending, LPD will finish spooling to JES any print jobs that it has received and is currently spooling. JES will handle these jobs after LPD ends.

This topic describes how to configure the LPD server. To operate the LPD server after it is running, see z/OS Communications Server: IP Configuration Reference.

The LPD server uses the Pascal socket API, so VMCF must be started for the server to successfully initialize. If VMCF is not started, message EZY1980E will be issued and the server will terminate.

Configuring the Remote Print Server

Procedure

Perform the following steps to configure the Remote Print Server:

1. Specify AUTOLOG and PORT statements in PROFILE.TCPIP.
2. Update the LPD server cataloged procedure.
3. Update the LPD server configuration data set.
4. Create a banner page (optional).

Results

For information about operating and controlling the LPD server, see z/OS Communications Server: IP Configuration Reference.

Step 1: Configuring PROFILE.TCPIP for LPD

If you want the LPD server started automatically when the TCPIP address space is started, include the name of the member containing the LPD server cataloged procedure in the AUTOLOG list in *hlq*.PROFILE.TCPIP. For example:

```
AUTOLOG
  LPSERVE
ENDAUTOLOG
```

To ensure that port TCP 515 is reserved for the LPD server, also add the name of the member containing the LPD cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP. For example:

```
PORT
  515 TCP LPSERVE
```

See the z/OS Communications Server: IP Configuration Reference for more information about the AUTOLOG and PORT statements.

© Copyright IBM Corp. 2000, 2022

Step 2: Updating the LPD server cataloged procedure

Update the LPD server cataloged procedure to suit your local configuration by copying the sample to your system or recognized PROCLIB from SEZAINST(LPSPROC) and modifying the sample. Specify LPD parameters, and change the data set names as required. See “Specifying LPD server parameters” on page 1270 for more information on the LPD server parameters.

Specifying LPD server parameters

The system parameters required by the LPD server are passed by the PARM option on the EXEC statement of the LPD cataloged procedure. Update the following parameters as required.

LPDDATA=‘*data_set_name*’
Specifies the fully qualified name of the data set containing the LPD configuration statements. This data set can be sequential or a member of a PDS.

LPDPRFX=‘PREFIX *your_prefix*’
Specifies the high-level qualifier to be used for temporary data sets created by the LPD server. Include both the PREFIX keyword and your qualifier in the quoted string. The qualifier may be up to 26 characters. If it is blank, the default is the procedure name. The LPD task requires the authority to create and modify data sets with this prefix.

DIAG=‘*options*’
Specifies any of the following diagnostic options in a quoted string of keywords separated by blanks. For example, `DIAG='VERSION TRACE'`

VERSION
Displays the version number.

TYPE
Activates high-level trace facility in the LPD server. Significant events, such as the receipt of a job for printing, are recorded in the //SYSOUT DD data set specified in your LPD server cataloged procedure.

TRACE
Causes a detailed trace of activities within the LPD server to record in the //SYSOUT DD data set specified in your LPD server cataloged procedure. The detailed tracing can also be activated with the DEBUG statement in the LPD server configuration data set and with the TRACE command of the SMSG interface.

Note: The JCL PARM= statement has a limit of 100 characters.

Configuring LPDDATA

Use a DD card that requires the LPD configuration file to be in a data set.

```
      //SYSLPDD  DD DISP=SHR,DSN=TCPIP.SEZAINST(LPDDATA)
```

Using this example, the DD card must be specified as SYSLPDD, but the data set name can be any valid data set name with a member specified up to 44 characters.

To use the DD card method, you must comment out or remove the LPDDATA= parameter from the PROC statement and remove the "&LPDDATA'' PARM from the LPD EXEC statement.

Note: The search order for the configuration file is:

1. LPDDATA= on the PARM= statement
2. //SYSLPDD DD statement
3. *hlq*.LPD.CONFIG

If both the LPDDATA= statement on the PARM= statement and the //SYSLPDD DD statement are specified, the data set name specified on LPDDATA= is used.

The LPD server does not limit the number of print jobs it handles per connection. This can cause a memory abend to occur if too many print jobs are sent in one connection. Certain LPR clients, such as SUN UNIX, are set up to send multiple jobs in one connection. It is recommended that the LPD start procedure be started with a region size of 6M and the LPR client send no more than 50 print jobs in one connection.

Step 3: Updating the LPD server configuration data set

The LPD configuration data set defines the local, NJE, and remote services (printers and punches) used by the LPD server. To update the LPD server configuration data set, copy the sample provided in SEZAINST(LPDDATA) and modify it to suit your installation. See z/OS Communications Server: IP Configuration Reference for more details.

- To define a printer or punch:
  - Include a SERVICE statement with the appropriate parameters for each printer or punch your are defining.
  - Specify the type of service with the LOCAL, NJE, or REMOTE parameter in the SERVICE statement.
  - For local or NJE services, include any of the optional parameters to further define the service: EXIT, FAILEDJOB, FILTERS, LINESIZE, PAGESIZE and RACF.
  - For remote services, specify the destination printer and host. Any additional specifications are defined on the remote system and are not necessary in the SERVICE statement.
- To turn on LPD server tracing, include the optional DEBUG statement.
- To authorize users for the SMSG interface, include the optional OBEY statement.
- Printer names cannot contain an at sign (@).

Step 4: Creating a banner page (optional)

LPBANNER is the name of the default program that is provided in executable form in the SEZATCP data set. This program is specified on the EXIT parameter in the SERVICE statement. LPBANNER prints a separator page that contains, in large letters, a banner stating “LPD BANNER”, the user ID, the job name, and the job class. Field headings of HOST, USER, JOB, and CLASS appear in smaller letters.

The sample exit LPBANNER uses machine carriage control and is designed to be used with the SERVICE PRINTER LOCAL or SERVICE PRINTER NJE statements. To use this sample exit, both SERVICE statements require a printed LINESIZE of 78 or greater. Banner pages are usually not used with the SERVICE PUNCH or SERVICE RECFMU statements; however, if you want to have banner pages (headers) for the SERVICE PUNCH or SERVICE RECFMU devices, a user created banner exit is required.

You can either use the executable form or copy and modify the sample source provided in SEZAINST(EZAAE04S) and SEZAINST(EZAAE04T) to create a banner. If you are changing the source to create your own banner, assemble and link-edit these data sets as reentrant. You can modify and use the following JCL to do this. Changing the ENTRY and NAME to something other than LPBANNER will avoid possible maintenance problems in the future.

```
//ASMLNK JOB MSGLEVEL=(1,1),MSGCLASS=A,CLASS=A,REGION=1024K
//ASM1   EXEC PGM=ASMA90,PARM='OBJECT,XREF(FULL)'
//STEPLIB DD  DISP=SHR,DSN=HLA.OSV1R4.SASMMOD1
//*  ASSEMBLER H
//SYSLIB   DD DSN=tcpip_hlq.SEZACMAC,DISP=SHR
//         DD DSN=SYS1.MACLIB,DISP=SHR
//         DD DSN=SYS1.MODGEN,DISP=SHR
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(5,5)),DSN=&&SYSUT1
//SYSPUNCH DD DUMMY,DCB=BLKSIZE=80
//SYSPRINT DD SYSOUT=A
//SYSLIN   DD DSN=&&OBJECT(EZAAE04S),DISP=(,PASS),UNIT=SYSDA,
//            SPACE=(CYL,(5,5,1)),DCB=BLKSIZE=400
//SYSIN    DD DSN=tcpip_hlq.SEZAINST(EZAAE04S),DISP=SHR
/*
//ASM2   EXEC PGM=ASMA90,PARM='OBJECT,NODECK,XREF'
//STEPLIB  DD  DISP=SHR,DSN=HLA.OSV1R4.SASMMOD1
//*  ASSEMBLER H
//SYSLIB   DD DSN=tcpip_hlq.SEZACMAC,DISP=SHR
```

```
//         DD DSN=SYS1.MACLIB,DISP=SHR
//         DD DSN=SYS1.MODGEN,DISP=SHR
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(5,5)),DSN=&&SYSUT1
//SYSPUNCH DD DUMMY,DCB=BLKSIZE=80
//SYSPRINT DD SYSOUT=A
//SYSLIN   DD DSN=&&OBJECT(EZAAE04T),DISP=(OLD,PASS)
//SYSIN    DD DSN=tcpip_hlq.SEZAINST(EZAAE04T),DISP=SHR
/*
//LNK    EXEC PGM=IEWL,PARM='LIST,NCAL,RENT,LET'
//SYSPRINT DD SYSOUT=A
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(1,1))
//SYSLMOD  DD DSN=tcpip_hlq.SEZATCP,DISP=SHR
//AEZAMODS DD DSN=tcpip_hlq.AEZAMODS,DISP=SHR
//OBJECT   DD DSN=&&OBJECT,DISP=(OLD,DELETE)
//SYSLIN   DD *
  ORDER   EZBOECPR
  INCLUDE AEZAMODS(EZBOECPR)
  INCLUDE OBJECT(EZAAE04S)
  INCLUDE OBJECT(EZAAE04T)
  MODE    AMODE(24),RMODE(24)
  ENTRY   LPBANNER
  NAME    LPBANNER(R)
/*
```

Chapter 25. Remote procedure calls

This topic contains information to help you configure the following applications and interface:

- PORTMAP
- UNIX PORTMAP
- rpcbind
- NCS

Restriction: PORTMAP, UNIX PORTMAP, and rpcbind cannot be run concurrently, because all three applications listen on the well known /etc/services sunrpc port of 111. You must determine which server will service your RPC applications.

For information about rpcbind configuration and setup in a multilevel secure environment, see "z/OS UNIX rpcbind server" on page 206.

Read "Understanding search orders of configuration information" on page 13. It covers important information about data set naming and search sequences.

Steps for configuring the PORTMAP address space

The PORTMAP address space runs the Portmapper function. Portmapper is the software that supplies client programs with the port numbers of server programs.

About this task

Clients contact server programs by sending messages to port numbers where receiving processes receive the message. Because you make requests to the port number of a server rather than directly to a server program, client programs need a way to find the port number of the server programs that they want to call. Portmapper standardizes the way clients locate the port number of the server programs supported on a network.

Procedure

Perform the following steps to configure PORTMAP:

1. Specify AUTOLOG and PORT statements in *hlq*.PROFILE.TCPIP.
2. Configure security server (or RACF equivalent) items.
3. Update the PORTMAP cataloged procedure.
4. Define the data set for well-known procedure names.

Step 1: Configuring PROFILE.TCPIP for PORTMAP

If you want the PORTMAP server to start automatically when the TCPIP address space is started, you should include PORTMAP in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

```
AUTOLOG
  PORTMAP
ENDAUTOLOG
```

Note: If your system is using the Network File System (NFS) server, you must start the PORTMAP address space. See z/OS Network File System Guide and Reference for more information.

© Copyright IBM Corp. 2000, 2022

To ensure that port UDP 111 and TCP 111 are reserved for the PORTMAP server, also add the name of the member containing the PORTMAP cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP:

```
PORT
  111 UDP PORTMAP
  111 TCP PORTMAP
```

See the z/OS Communications Server: IP Configuration Reference for more information about the AUTOLOG and PORT statements.

Step 2: Configuring security server (or RACF equivalent) items

The PORTMAP cataloged procedure assumes that the procedure has the authority to run as a started task. To ensure that the PORTMAP procedure has the appropriate security server access, enter the following commands as shown in SEZAINST(EZARACF):

```
ADDUSER  PORTMAP DFLTGRP(OMVSGRP) OMVS(UID(0)
HOME('/'))NOPASSWORD
RDEFINE  STARTED PORTMAP.* STDATA(USER(PORTMAP))
SETROPTS RACLIST(STARTED) REFRESH
SETROPTS GENERIC(STARTED) REFRESH
```

Step 3: Updating the PORTMAP cataloged procedure

Update the PORTMAP cataloged procedure to suit your local conditions by copying the sample provided in SEZAINST(PORTPROC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Change the data set names as required. See z/OS Communications Server: IP Configuration Reference for more details.

Step 4: Defining the data set for well-known procedure names

z/OS Communications Server includes a data set that contains a list of commonly used procedure names. This data set is used by the RPCINFO command to resolve remote program numbers into their equivalent program names.

To create the data set, copy SEZAINST(ETCRPC) to the default data set called *hlq*.ETC.RPC. If a user has a *user_id*.ETC.RPC data set defined, it takes precedence over the preceding data set.

Normally, you would not change this data set except to add a new application to the list. To add a new application, add a line that contains the following items:

- The *program_name* of the new application or procedure
- The *program_number* of the new application or procedure
- Any comments regarding the description of the program

The items are variable in format, each separated by a blank.

The SEZAINST(ETCRPC) data set contains the well-known procedure names. The following sample is the ETCRPC sample.

```
#  z/OS Communications Server
#  SMP/E distribution path: /usr/lpp/tcpip/samples/IBM/EZAOERPC
#  SMP/E distribution path: SEZAINST(EZAEB01X)
#
 #  Licensed Materials - Property of IBM #  5650-ZOS Copyright IBM Corp. 1996, 2015 #  Status =
CSV2R2
#
#  Change Activity:
#
#  Flag Reason   Release  Date   Origin    Description
#  ---- -------- -------- ------ --------  -----------------------
#  $Y1= D139012  R9BASEN  061011 staff    : NFS information added
#  $F1= RBAPPREM CSV1R11  080728 staff    : Remove NDB

#  $31= RFSLEDLC CSV2R2   140731 adamson  : Remove x25.inr (23281)
```

```
# -----------------------------------------------------------------
# Name      ProgNumber  Alias names, if any
#
portmapper    100000    portmap sunrpc
rstatd        100001    rstat rup perfmeter
rusersd       100002    rusers
nfsd          100003    nfs nfsprog         # Network File System daemon
ypserv        100004    ypprog
mountd        100005    mount showmount     # Mount daemon
ypbind        100007
walld         100008    rwall shutdown
yppasswdd     100009    yppasswd
etherstatd    100010    etherstat
rquotad       100011    rquotaprog quota rquota
sprayd        100012    spray
3270_mapper   100013
rje_mapper    100014
selection_svc 100015    selnsvc
database_svc  100016
rexd          100017    rex
alis          100018
sched         100019
llockmgr      100020
nlockmgr      100021    nlm nfs_lockd        # Network Lock Manager

statmon       100023

status        100024    nsm nfs_statd        # Network Status Monitor
mvsmount      100044    nfs_mvsmnt           # MVSmount daemon
showattr      100059    nfs_showattr         # showattr daemon
pcnfsd        150001    nfs_pcnfs            # pcnfs daemon
```

Starting the PORTMAP address space

If you did not include PORTMAP in the AUTOLOG statement, you can start PORTMAP with the START command. For example:

```
    START PORTMAP
```

Steps for configuring the z/OS UNIX PORTMAP address space

The z/OS UNIX PORTMAP address space runs the Portmapper function.

Procedure

Perform the following steps to configure z/OS UNIX PORTMAP:

1. Specify PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the PORTMAP cataloged procedure.

Results

Step 1: Configuring PROFILE.TCPIP for UNIX PORTMAP

If you want the PORTMAP server to start automatically when the TCPIP address space is started, you should include PORTMAP in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

```
AUTOLOG
  PORTMAP JOBNAME PORTMAP1
ENDAUTOLOG
```

To ensure that port UDP 111 and TCP 111 are reserved for the z/OS UNIX PORTMAP server, add the z/OS UNIX PORTMAP server jobname to the PORT statement in *hlq*.PROFILE.TCPIP. If you use the sample cataloged procedure, PORTMAP, to start the z/OS UNIX PORTMAP server, the jobname is PORTMAP1:

```
PORT
  111 UDP PORTMAP1     ; Portmapper Server
  111 TCP PORTMAP1     ; Portmapper Server
```

See the z/OS Communications Server: IP Configuration Reference for more information about the PORT statement.

Step 2: Updating the PORTMAP cataloged procedure

Update the PORTMAP cataloged procedure to suit your local conditions by copying the sample provided in SEZAINST(OPORTPRC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Change the data set names as required. See z/OS Communications Server: IP Configuration Reference for more details.

Starting the PORTMAP address space

There are two ways to start the portmapper as a z/OS UNIX socket application:

- From the z/OS shell
- As a started task

To start the portmapper from the z/OS shell, the user ID must be an authorized superuser. The authorized superuser ID can issue **oportmap &** to start the portmapper. For the authorization procedure, see z/OS UNIX System Services Planning.

You can also start PORTMAP as a started task with the START command as follows:

```
    START PORTMAP
```

Note: If your system is using the Network File System (NFS) server, see z/OS Network File System Guide and Reference for more information.

Steps for configuring the rpcbind address space

The rpcbind software supplies client programs with the universal addresses of RPC server programs.

Before you begin

The z/OS rpcbind server supports the portmapper binding protocol (RPC Binding Protocol Version 2) as well as other RPC binding protocols. This means you can run rpcbind instead of portmapper. Portmapper and rpcbind cannot be run at the same time, because they both listen on /etc/services sunrpc port 111.

About this task

RPC server applications do not have well-known port numbers. Instead, they obtain an ephemeral port number, create a universal address, and register that address with rpcbind. Clients contact server programs by obtaining the universal address of the server from rpcbind, and sending messages to the universal address of the server.

Procedure

Perform the following steps to configure the rpcbind address space.

1. Configure the PROFILE.TCPIP data set for rpcbind.
2. Configure security server (or RACF equivalent) items.
3. Update the RPCBIND cataloged procedure.
4. Update the /etc/services file.

5. Configure SYS1.PARMLIB for rpcbind.

Step 1: Configuring the PROFILE.TCPIP data set for rpcbind

Guideline: The rpcbind server is a generic server, as described in "Generic servers in a CINET environment" on page 48.

- In a single stack environment, you can use the AUTOLOG statement to automatically start generic servers.
- In a multiple stack environment, it is not advisable to use the AUTOLOG statement to start generic servers. You can use an operations automation software package that IBM or other companies provide to automatically start generic servers.

If you want the rpcbind server to start automatically when the TCPIP address space is started, include rpcbind in the AUTOLOG statement in the PROFILE.TCPIP data set.

```
AUTOLOG
  RPCBIND JOBNAME RPCBIND1
ENDAUTOLOG
```

To ensure that port UDP 111 and TCP 111 is reserved for the rpcbind server, add the rpcbind server job name to the PORT statement in the PROFILE.TCPIP data set. If you use the sample cataloged procedure, RPCBIND, to start the rpcbind server, the job name is RPCBIND1.

```
PORT
  111 UDP RPCBIND1     ; rpcbind server
  111 TCP RPCBIND1     ; rpcbind server
```

If you start the rpcbind server from the z/OS UNIX shell, the job name is OMVS.

```
PORT
  111 UDP OMVS         ; rpcbind server OMVS
  111 TCP OMVS         ; rpcbind server OMVS
```

For more information about the PORT statement, see z/OS Communications Server: IP Configuration Reference.

Step 2: Configuring security server (or RACF equivalent) items

The RPCBIND cataloged procedure assumes that the procedure has the authority to run as a started task. To ensure that the RPCBIND procedure has the appropriate security server access, enter the following commands as shown in SEZAINST(EZARACF):

```
ADDUSER RPCBIND DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(0) HOME('/'))
RDEFINE STARTED RPCBIND.* STDATA(USER(RPCBIND))
SETROPTS RACLIST(STARTED) REFRESH
SETROPTS GENERIC(STARTED) REFRESH
```

You can define the SAF resource profile EZB.RPCBIND.*sysname.rpcbindname*.REGISTRY in the SERVAUTH class to control which users can register or deregister applications with rpcbind. You can use wildcards. For example, if you use wildcard values for *sysname* and *rpcbindname*, the profile name is as follows:

```
EZB.RPCBIND.*.*.REGISTRY
```

In this example, suppose the MVS system name is MVS000 and the RPCBIND catalogued procedure is used to start the rpcbind server. This procedure uses the job name RPCBIND. RPCBIND is fewer than 8 characters, so the *rpcbindname* is RPCBIND1, and the profile name is as follows:

```
EZB.RPCBIND.MVS000.RPCBIND1.REGISTRY
```

The profile EZB.RPCBIND.*sysname.rpcbindname*.REGISTRY is optional. If it is not defined, all users can register and deregister applications with rpcbind. If the profile is defined, only users granted at least READ access to this resource profile can register or deregister applications with rpcbind.

In this example, if your SAF security product is RACF and you want only the RPC server TRUESERV running under user ID TRUESERV to be able to register and deregister applications with rpcbind, you can use the following commands to define the profile EZB.RPCBIND.*.*.REGISTRY in the SERVAUTH class and grant TRUESERV read access to the profile:

```
RDEFINE SERVAUTH EZB.RPCBIND.*.*.REGISTRY UACC(NONE)
PERMIT EZB.RPCBIND.*.*.REGISTRY ID(TRUESERV) ACCESS(READ) CLASS(SERVAUTH)
```

Requirements for a multilevel secure environment:

- The profile EZB.RPCBIND.*sysname.rpcbindname*.REGISTRY is mandatory, and you must grant user IDs associated with trusted RPC servers at least READ access to this profile. For more information about running rpcbind in a multilevel secure environment, see “z/OS UNIX rpcbind server” on page 206.
- If the SAF FACILITY class resource profile BPX.POE is defined, and your installation directs target assistance calls to rpcbind, you must grant the user ID assigned to rpcbind at least READ access to this profile. For more information on target assistance, see z/OS Communications Server: IP Programmer's Guide and Reference.

Tips:

- The target assistance RPCs PMAPPROC_CALLIT, RPCBPROC_CALLIT, RPCBPROC_BCAST, and RPCBPROC_INDIRECT are defined in RFC 1833 (*Binding Protocols for ONC RPC Version 2*).
- The RPC library routines clnt_broadcast and pmap_rmtcall() issue a target assistance request on behalf of the caller.
- The rpcinfo utility issues a target assistance call when it is invoked with the -b option.

Guideline: If RPCBIND is the user ID assigned to the rpcbind server, you can use the following command to grant the user ID READ access to the profile:

```
PERMIT   BPX.POE CLASS(FACILITY) ID(RPCBIND) ACCESS(READ)
```

Step 3: Updating the RPCBIND cataloged procedure

Update the RPCBIND cataloged procedure to suit your local conditions by copying the sample provided in SEZAINST(RPCBIND) to your system or recognized PROCLIB, and modifying it as necessary. Change the data set names as required. For further information, see “Starting the rpcbind address space” on page 1279. A copy of RPCBIND is also available in z/OS Communications Server: IP Configuration Reference.

Step 4: Updating the /etc/services file

Update /etc/services to ensure that the sunrpc service is reserved for TCP and UDP port 111:

```
sunrpc    111/tcp
sunrpc    111/udp
```

Restriction: If you change the sunrpc port, rpcinfo on that host will not be able to contact the rpcbind server.

Step 5: Configure SYS1.PARMLIB for rpcbind

The rpcbind server uses semaphore sets and shared memory segments. Inspect your BPXPRM*xx* member in SYS1.PARMLIB and verify that the following statements are configured such that rpcbind will be able to obtain three shared memory segments of one page each, and three semaphore sets:

```
IPCSHMMPAGES
IPCSHMNSEGS
IPCSEMNIDS
IPCSEMNSEMS
```

For more information about these SYS1.PARMLIB statements, see z/OS MVS Initialization and Tuning Reference.

Tip: You can issue the operator command to display these values.

Starting the rpcbind address space

There are two ways to start rpcbind as a z/OS UNIX socket application:

- From the z/OS shell
- As a started task

It cannot be started from TSO.

To start rpcbind from the z/OS shell, the user ID must be an authorized superuser. The authorized superuser ID can issue to start rpcbind. For information about assigning superuser attributes, see z/OS UNIX System Services Planning.

You can also start rpcbind as a started task from the MVS console with the START command, as follows:

```
START RPCBIND
```

You can specify the following options when you are starting rpcbind:

- Specify `-d` for rpcbind to send trace information to the daemon facility of syslogd.
  - `-df` sends flow information.
  - `-dl` sends log information of all RPC procedures called.
  - `-dx` sends XDR information.
- The `-i` option enables you to specify the z/OS UNIX file system directory to which the pid file is written. The pid file name is always rpcbind.pid. If `-i` is not specified, the rpcbind process ID is written to /etc/rpcbind.pid.

  Restriction: The directory path name must be an absolute path name. That is, it must begin with the forward slash (/) character, as shown in the following example:

  ```
  rpcbind -i /tmp
  ```
- The `-n` option enables you to direct rpcbind to run in a non-swappable environment. A process might need to run as non-swappable to ensure that it is available during periods of high CPU usage. However, a non-swappable process might convert real storage in the system to preferred storage. Because preferred storage cannot be configured offline, allowing rpcbind to run in a non-swappable state can reduce the future ability of your installation to reconfigure storage.

  If you do specify the `-n` option, ensure that the user ID associated with rpcbind has at least READ access to the resource profile BPX.STOR.SWAP in the FACILITY class. The default is to start rpcbind as swappable.
- The `-s` option specifies the number of statistics entries per binding protocol that rpcbind maintains. Valid values are in the range 113 – 500. Statistics that are maintained by the rpcbind server are used to reply to the RPCBPROC_GETSTAT request. For more information about statistics that are maintained by the rpcbind server, see RFC 1833.

  Result: Rpcbind calculates the number of pages that are needed to store statistics for the value that is specified, and obtains that number of pages of shared memory for statistics. Rpcbind rounds up the number of statistics entries it tracks to fully use the shared memory.

  Tip: Rpcbind does not start unless it can obtain sufficient shared memory to maintain statistics for the number of entries you specify. You configure the number of pages of shared memory available to z/OS with the IPCSSHMMPAGES parameter in the BPXPRM*xx* member of SYS1.PARMLIB.
- To display help information, specify the `-?` option.

Tip:

- If your system is using the Network File System (NFS) server, for more information about initialization attributes for the z/OS client and Network File System operation, see z/OS Network File System Guide and Reference.
- The rpcbind server uses ENF event code 80 to send signals when rpcbind completes initialization and when rpcbind is stopping. For more information, see Using ENF event code 80 to listen for rpcbind events in z/OS Communications Server: IP Programmer's Guide and Reference.

Restriction: Portmapper and rpcbind cannot be run at the same time, as they both listen on `/etc/services` sunrpc port 111.

Steps for configuring the NCS interface

This topic describes how to configure the Network Computing System (NCS).

NCS is the Remote Procedure Call (RPC) implementation of Apollo's Network Computing Architecture (NCA**).

NCS includes:

1. RPC runtime library
2. Location broker
3. Network Interface Definition Language (NIDL) compiler

The RPC runtime library and the location broker provide runtime support. Together these two elements make up the Network Computing Kernel (NCK) which includes all the software necessary to run a distributed application. The NIDL compiler is a tool for developing NCS-based applications.

In order to execute NCS applications in an MVS environment, you must start a local location broker (LLBD). One of the hosts in your IP network must also start the global location broker (GLBD). Both the LLBD and the NRGLBD maintain information about active NCS server applications.

- LLBD server

  The LLBD manages the LLB database which stores information for the NCS-based servers running on this host.

  Your host must run LLBD if you want to support the location broker forwarding function or allow remote access to the LLB database. The LLBD function must be started on the host before any other NCS-based servers are started and before the NRGLBD is started.

  The LLB database is stored in the data set ADM@SRV.LLB@LL.DATABASE, which is not created until an entry is registered with the LLBD. This data set can be administered using the lb@admin tool.
- NRGLBD server

  The NRGLBD manages the NCS global location broker (GLB) database. The GLB helps clients locate servers on a network or internet.

  There are two versions of the GLB daemon: replicated GLBD and NRGLBD. The replicated GLBD is available only on Apollo, SunOS, and Ultrix systems. For other systems, such as IBM, only the NRGLBD is available. The advantage of replicated GLBD over NRGLBD is that the GLBD can be run at the same time on several network hosts, providing greater availability in the event that one of the hosts running GLBD fails or if there is a partial network failure.

  You cannot use the NRGLBD on your system if:

  – The replicated version of GLBD can run on some other host in your network
  – Another host in your network is already running NRGLBD

  The GLB database is stored in a data set ADM@SRV.LLB@LG.DATABASE. This data set is not created until an entry is registered with the NRGLBD.

  The record structure for the LLBD and the NRGLBD databases is identical.

Perform the following steps to configure NCS:

1. Specify AUTOLOG and PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the NRGLBD cataloged procedure in SEZAINST(NRGLBD).
3. Update the LLBD cataloged procedure in SEZAINST(LLBD).

For more information about NCS, see *Network Computing System Reference Manual*.

Step 1: Configuring PROFILE.TCPIP for NCS

If you want LLBD and NRGLBD to start automatically when the TCPIP address space is started, then you should include the names of the members containing the LLBD and NRGLBD cataloged procedures in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

The LLBD must be running before you start NRGLBD. Therefore, you must put LLBD before NRGLBD in the AUTOLOG statement.

```
AUTOLOG
  LLBD
  NRGLBD
ENDAUTOLOG
```

To ensure that port UDP 135 is reserved for the LLBD server, add the name of the member containing the LLBD cataloged procedure to the PORT statement in the *hlq*.PROFILE.TCPIP data set.

```
PORT
  135 UDP LLBD
```

See the z/OS Communications Server: IP Configuration Reference for more information about the AUTOLOG and PORT statements.

Step 2: Updating the NRGLBD cataloged procedure

Update the NRGLBD cataloged procedure by copying the sample provided in SEZAINST(NRGLBD) to your system or recognized PROCLIB and modifying it to suit your local conditions. See the NCS topic in z/OS Communications Server: IP Configuration Reference for more details.

Step 3: Updating the LLBD cataloged procedure

Update the LLBD cataloged procedure by copying the sample provided in SEZAINST(LLBD) to your system or recognized PROCLIB and modifying it to suit your local conditions. See the NCS topic in z/OS Communications Server: IP Configuration Reference for more information.

Chapter 26. Mail on z/OS

This topic describes how to configure the Communications Server SMTP (CSSMTP) application and popper.

CSSMTP is the strategic mail program for z/OS Communications Server and should be used for forwarding mail from your NJE network or from local applications that write mail to a JES spool file, using the existing SMTP batch interface to forward the mail to other gateway servers. CSSMTP provides the easiest and simplest solution for this purpose.

The sendmail to CSSMTP bridge is a replacement for the z/OS UNIX sendmail command for sending mail:

- The sendmail bridge processes the sendmail command input options (supported command line switches and configuration statements), reads the mail file, and constructs a CSSMTP compatible file with SMTP commands.
- Basic sendmail commands that use the sendmail bridge should function seamlessly, but more advanced options might require alternate solutions.
- The sendmail bridge requires CSSMTP to be configured and running.

Note: CSSMTP and the sendmail bridge don't provide support for receiving mail for delivery to local TSO/E or z/OS UNIX System Services user mailboxes, or for forwarding mail to other destinations

Configuring the CSSMTP application

Communications Server Simple Mail Transfer Protocol (CSSMTP) is a mail-forwarding SMTP client application. CSSMTP processes data sets containing mail messages on the spool file and forwards them to a target message transfer agent (MTA) without resolving each recipient. CSSMTP can improve the performance, scalability, and availability of the client function, but it does not act as a listening MTA server. Multiple instances of CSSMTP can run on a single host.Figure 156 on page 1283 shows how CSSMTP fits into a network.

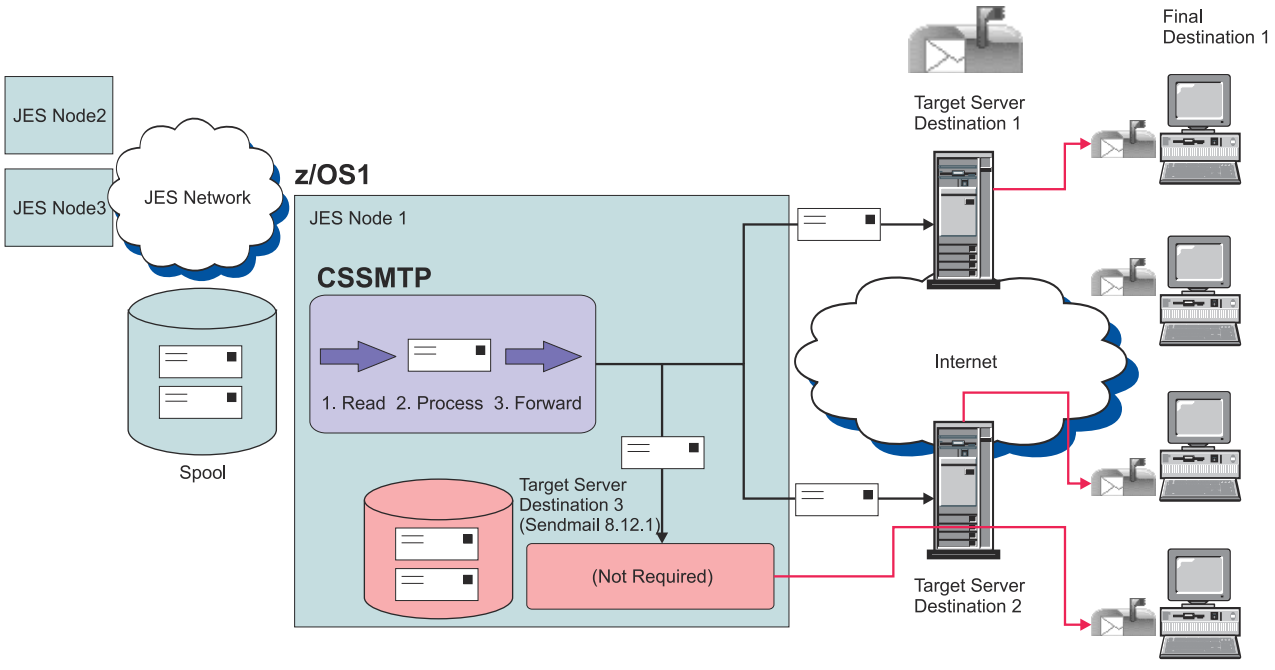


Figure 156. CSSMTP forwards mail messages from spool to the network

CSSMTP implements RFC 2821 and RFC 2822 for interacting with server MTAs, and supports additional RFCs for message size (RFC 1870) and security (RFC 3207). CSSMTP is not a fully capable MTA and

© Copyright IBM Corp. 2000, 2022

functions as an outbound forwarder, sending mail messages from the JES spool data set to the Internet. As the mail messages are read from the spool file, CSSMTP functions like a TCP/IP SMTP client and interacts with one or more configured target servers. When processing mail messages from the spool file, CSSMTP does not resolve mail message recipients, but transfers mail messages to one or more configured, next-hop servers (fully capable MTAs) that might or might not be the final destination.

CSSMTP provides the following capabilities:

- Checkpoint capabilities

  If you need to restart CSSMTP, it does not have to reprocess the entire spool file, which will reduce duplicate mail that is received by message recipients.
- Long retry and extended retry capabilities
  - The long retry capability is available for up to 5 days, and is intended to compensate for short-term target server outages.
  - The extended retry capability is available for a longer period or for an indefinite period, and is intended to compensate for extended target server outages.

  For more information, see "Customizing the CSSMTP configuration file to try mail again" on page 1289.
- Multiple security capabilities

  See "Security for CSSMTP" on page 1292.
- SMF recording of CSSMTP events

  See "Steps for configuring SMF records for CSSMTP (optional)" on page 1295.

As stated in RFC 2821, SMTP clients that transfer all traffic, regardless of the target domain names that are associated with the individual mail messages, or that do not maintain queues for trying mail message transmissions again that initially cannot be completed, might otherwise conform to this specification but are not considered fully capable. CSSMTP does not implement all aspects of RFC 2821.

Terms and concepts

The following terms and concepts are used in this information:

Backpressure
CSSMTP stops processing the JES spool file when mail cannot be sent to the target because the connection to a target server is down or unresponsive. CSSMTP resumes processing the JES spool file when a target server connection becomes active.

Bad spool file
A bad spool file is a JES spool file that cannot be processed as the result of security, configuration, syntax, readability, or delivery issues. The action taken by CSSMTP is subject to the setting of the BadSpoolDisp statement. For more information about the BadSpoolDisp statement, see z/OS Communications Server: IP Configuration Reference.

Checkpointing
If checkpointing is enabled using the CHKPOINT DD statement in the CSSMTP started procedure and there are partially processed spool files on the spool when restarting, CSSMTP attempts to send only the mail from these spool files that was not previously sent, which reduces the duplicate mail that is received by recipients. If CSSMTP is started cold using the `-f` option, any existing checkpoint records are flushed before CSSMTP is restarted. For information about using the CHKPOINT DD statement in the CSSMTP started procedure and about starting CSSMTP with the `-f` start option, see z/OS Communications Server: IP Configuration Reference.

Dead letter
A dead letter is created and stored for the following reasons:

- An undeliverable mail notification cannot be returned to the originator
- The original mail message was undeliverable, but did not specify an originator
- An error report cannot be delivered to the mail administrators

To ensure that dead letters are stored, configure Store on the DeadLetterAction parameter of the Undeliverable statement. For information about the Undeliverable statement, see z/OS Communications Server: IP Configuration Reference.

For examples of customizing the configuration to handle undeliverable mail, see "Customizing the CSSMTP configuration file to handle undeliverable mail" on page 1290.

ESMTP
Extended Simple Mail Transfer Protocol

Extended retry
If a mail message is not sent to all recipients because the target servers replied with a retry reply code, another attempt to send the message is made after the delay that is defined by the RetryLimit statement. If the long retry limits that are defined by the RetryLimit statement are exceeded, the mail is placed in the extended retry state. The mail is saved in a z/OS UNIX file system directory for extended retry as defined by the ExtendedRetry statement. If the extended retry limit is exceeded, the mail is marked as undeliverable. For more information about mail retry, see "Customizing the CSSMTP configuration file to try mail again" on page 1289. For information about using the RetryLimit statement and the ExtendedRetry statement to configure the retry limits, see z/OS Communications Server: IP Configuration Reference.

Long retry
See the definition for extended retry.

Mail administrator
The mail administrator is a special user ID for mailing systems, to which CSSMTP delivers error reports for problems detected while processing a spool file from the JES spool data set. For information about using the MailAdministrator statement to configure up to four mail administrator addresses, see z/OS Communications Server: IP Configuration Reference.

Message content
The headers and the structured body of a mail message. A separate Multipurpose Internet Mail Extensions (MIME) specification provides the definitions for structured bodies.

MTA
Message transfer agent

SMTP
Simple Mail Transfer Protocol

SMTP command
A command sent from the client to the SMTP server that lets the server know what information is being sent. For example, MAIL FROM: is a command.

SMTP reply
The acknowledgement, positive or negative, sent from an SMTP server. Replies are in the ASCII character set of the target server.

SMTP server
A network application implementing the SMTP protocol that accepts mail messages from SMTP clients.

Target server
A target server is the resolved or configured IP address from a TargetServer statement. The first four target servers are used by CSSMTP. For information about using the TargetServer statement to configure a target server, see z/OS Communications Server: IP Configuration Reference.

Undeliverable mail
Mail messages that cannot be delivered. CSSMTP processes and tries to send all mail messages in the spool file. Sometimes, one or more of these mail messages might be undeliverable. Examples of what causes undeliverable mail include:

- A problem with the mail message itself, such as the message size
- A problem with the recipient record (RCPT TO:), such as an unknown mailbox
- A mail message requires a secure connection and none of the target servers support the STARTTLS SMTP command

- Security problems
- A networking problem, such as being unable to reach a target host that has the capabilities that are needed for the mail message

For examples of customizing the configuration to handle undeliverable mail, see "Customizing the CSSMTP configuration file to handle undeliverable mail" on page 1290.

Undeliverable mail notification
CSSMTP creates an undeliverable mail notification when the ReturntoMailFrom parameter is set to YES on the Undeliverable statement and the mail message cannot be delivered. This notification contains the reply from the target server, describing the reason that the mail message was undeliverable and the text of the original mail message. If the undeliverable mail notification cannot be sent immediately, it is not tried again and is subject to the setting of the DeadLetterAction parameter of the Undeliverable statement.

For examples of customizing the configuration to handle undeliverable mail, see "Customizing the CSSMTP configuration file to handle undeliverable mail" on page 1290.

Setting up CSSMTP

To set up CSSMTP to process and forward your mail, you must take the following actions:

- Configure and start CSSMTP.

  See "Steps for configuring and starting CSSMTP" on page 1286.
- Create mail on the JES spool data set for CSSMTP.

  See "Steps for creating mail on the JES spool data set for CSSMTP" on page 1287.

Steps for configuring and starting CSSMTP

You can use CSSMTP to process and forward your mail. These steps provide the minimum information that you need to configure and start CSSMTP.

Procedure

Perform the following steps to configure and start CSSMTP:

1. Create a new data set member in your procedure library for the CSSMTP JCL.

   A sample job is in SEZAINST(CSSMTP). Because the CSSMTP JCL sample contains a JOB card, you should define the procedure library that you use in the IEFJOBS or IEFPDSI concatenation of the master JCL, or put the new data set member in a subsystem procedure library such as SYS1.PROCLIB.
2. Define explicit authority for all user IDs that you want to be able to start CSSMTP.

   See "Steps for granting authority to start CSSMTP" on page 1291.
3. Customize the CSSMTP configuration file and set up at least one valid target server IP address using the TargetServer statement.

   A sample CSSMTP configuration file is included in member CSSMTPCF in SEZAINST. A target server is defined as the IP address that is resolved from or configured on the TargetServer statement. For information about using the TargetServer statement to configure a target server, see z/OS Communications Server: IP Configuration Reference. For information about other configuration statements used by CSSMTP, see z/OS Communications Server: IP Configuration Reference. For information about handling undeliverable mail, see "Customizing the CSSMTP configuration file to handle undeliverable mail" on page 1290.

   CSSMTP configuration statements are processed during the initialization of CSSMTP or when you issue the MODIFY *procname*,REFRESH command.
4. Set up the resolver search order.

   Use the default search order unless there are special circumstances that require you to use unique parameters. For example, if the resolver has parameters that are used only when the resolver is called by CSSMTP, you need to define those unique parameters. Define the unique parameters in a data set

that is specified on the SYSTCPD DD statement in the CSSMTP procedure JCL. For more information about resolvers and resolver configuration files, see Chapter 13, “The resolver,” on page 729.

5. Optionally, set up additional security for CSSMTP.

   See “Security for CSSMTP” on page 1292.

6. Optionally, allocate and define a Virtual Storage Access Method (VSAM) linear data set for the checkpoint function.

   A sample job is in SEZAINST(CSSMTPVL).

7. Set up the timezone using the TZ environment variable.

   If you do not set up the timezone, the timezone is not shown in the `Received` header line, which is used to indicate that CSSMTP has picked up a mail message. The default time value is used if the DATE header is not specified in the mail message.

8. Start CSSMTP by issuing the following command, where *csproc* is the CSSMTP procedure member name:

   ```
   START csproc
   ```

Results

You know you are done when initialization is complete and you have successfully connected to the specified target server for processing mail, as indicated by the following messages:

```
EZD1802I  csproc INITIALIZATION COMPLETE FOR  extWrtName
EZD1821I  csproc ABLE TO USE TARGET SERVER  ipAddress
```

Steps for creating mail on the JES spool data set for CSSMTP

This topic provides the minimum information that you need to create mail that can be processed and forwarded by CSSMTP.

Before you begin

You need to know the external writer name of the CSSMTP application that you want to process your mail data set. If the external writer name is not configured for CSSMTP using the ExtWrtName statement, then the default is the job name. For information about configuring the external writer name using the ExtWrtName statement, see z/OS Communications Server: IP Configuration Reference.

For more information about creating mail using CSSMTP commands, see z/OS Communications Server: IP User's Guide and Commands.

Procedure

Perform the following steps to create mail on the JES spool data set for CSSMTP:

1. Set up JES so that CSSMTP can create, read, write, and purge data from the JES spool data set.

   See “Steps for initial setup for CSSMTP” on page 1288.

2. Set up the mail to conform to the standardized syntax for text messages that are sent across networks.

   Mail messages have an envelope and contents. Envelopes contain all necessary information to accomplish transmission and delivery of the mail message content. The fields in the envelope are in a standard format.

3. Configure code page support.

   - If multi-byte character support is required, configure the MBCS YES statement.
   - Configure the code page of the input spool files with the TRANSLATE statement. The spool file must be written by using a code page that is supported by z/OS Unicode Services. The default TRANSLATE code page is IBM-1047. If MBCS YES is configured, the TRANSLATE code page must be a multi-byte character set code page.

- Configure the code page used by the mail servers with the Charset parameter defined on the TargetServer statements. The default Charset code page is ISO8859-1.
- If MBCS YES is configured, configure the code page used by the mail servers with the MBCharset parameter defined on the TargetServer statements. The MBCharset code page must be a multi-byte character set code page.

The spool file that contains the mail commands and body is recognized and translated by the Language Environment iconv function to the following code pages:

- The mail commands are translated from the TRANSLATE code page to EBCDIC (IBM-1047) for inspection by CSSMTP and then from IBM-1047 to the Charset code page before sending to the target server.
- The mail headers are translated from the TRANSLATE code page to EBCDIC (IBM-1047) for inspection by CSSMTP and then from the TRANSLATE code page to the Charset or MBCharset code page before sending to the target server.
- The mail body is translated from the TRANSLATE code page to the Charset or MBCharset code page before sending to the target server.

For information about Unicode Services, see z/OS Unicode Services User's Guide and Reference. For information about configuring the code page by using the MBCS statement, the Translate statement, and the Charset and MBCharset parameters on the TargetServer statement, see z/OS Communications Server: IP Configuration Reference.

4. Set up the mail in one of the following formats:
   - A flat file generated using the sendmail bridge command that writes to the SYSOUT data set

     For information about using the sendmail bridge command, see Sendmail to CSSMTP bridge in z/OS Communications Server: IP User's Guide and Commands.
   - A flat file generated using the IEBGENER utility or batch jobs that write to the SYSOUT data set

     For information about using the IEBGENER utility to copy a mail file to a JES SYSOUT data set, see z/OS Communications Server: IP User's Guide and Commands.
   - Netdata generated from SMTPNOTE or from TSO xmit

     For information about using the SMTPNOTE command to compose a single mail message to one or more recipients and about using the TSO TRANSMIT (XMIT) command to send a mail file, see z/OS Communications Server: IP User's Guide and Commands.

   You must copy and customize the SMTPNOTE CLIST on every system where users can send mail with the SMTPNOTE command. For more information, see “Steps for customizing the SMTPNOTE CLIST (optional)” on page 1289.

Results

You know you are done when mail is on the JES spool data set ready to be processed by CSSMTP.

Steps for initial setup for CSSMTP

Set up JES so that CSSMTP can create, read, write, and purge data from the JES spool data set.

Procedure

Perform the following steps to set up JES:

1. Set up JES2 or JES3 initialization parameters so that CSSMTP can interface with JES utilities to create, read, write, and purge data from the JES spool data set.

   For information about JES2 initialization, see z/OS JES2 Initialization and Tuning Guide. For information about JES3 initialization, see z/OS JES3 Initialization and Tuning Guide. For information about JES RACF authorization, see z/OS Security Server RACF Security Administrator's Guide.
2. Verify that JES exit programs do not interfere with the function of CSSMTP.

For JES2 exit and exit routine association, see z/OS JES2 Initialization and Tuning Reference. For JES3, if you are using the IATUX18 exit, see z/OS JES3 Customization.

3. Set up security authorization for the CSSMTP started task name in your definitions of authorized users.

   If you are using RACF for security, see z/OS Security Server RACF Security Administrator's Guide for information about RACF controls for the spool file.

4. Set up CSSMTP for warm starts by configuring the CHKPOINT DD statement.

   Checkpointing enables CSSMTP to recognize partially processed spool files, so that CSSMTP does not reprocess the entire spool file when restarting. This reduces the duplicate mail messages that are received by mail message recipients. Warm starts work only when restarting CSSMTP with the same job name and external writer name. If CSSMTP is started cold using the `-f` option, any existing checkpoint records are flushed before CSSMTP is restarted. For information about using the CHKPOINT DD statement in the CSSMTP started procedure and about starting CSSMTP, see z/OS Communications Server: IP Configuration Reference.

Steps for customizing the SMTPNOTE CLIST (optional)

Perform these steps for every system from which mail can be sent using the SMTPNOTE command.

Before you begin

If the ExtWrtName statement is specified in the CSSMTP configuration file, you should change the configured SMTPJOB value in the SMTPNOTE CLIST to match the ExtWrtName name of CSSMTP. If the ExtWrtName statement is not specified, you should customize the SMTPJOB name to match the CSSMTP job name. The SMTPJOB name must not be defined as a node name to JES and cannot begin with the characters R, RM, or RMT, because SMTPNOTE uses TSO XMIT to transmit the mail to CSSMTP.

Procedure

Perform the following steps to customize the SMTPNOTE CLIST:

1. Copy SEZAINST(SMTPNOTE) into the system CLIST data set.
2. Customize the variables in the SMTPNOTE CLIST, including the SMTPJOB variable so that it uses the CSSMTP job name. DOMAIN should also be set when using CSSMTP. For more information, see *Step 3: Customize the SMTPNOTE CLIST and modify parmlib data sets*.

Customizing the CSSMTP configuration file to try mail again

When mail servers reply to commands that resources are temporally unavailable, CSSMTP can try sending mail messages again. CSSMTP implements the following two types of retry:

- Long retry, defined by the RetryLimit statement

  When you configure long retry, mail messages are saved in CSSMTP memory while they await retry. The originating spool file is also held by CSSMTP until every mail message is delivered or until an undeliverable notice is sent. Because the originating spool file is held during retry processing, you should configure the RetryLimit statement to a time span that is shorter than the time span that you configure for extended retry. For example, the time span on the RetryLimit statement might be several hours.

- Extended retry, defined by the ExtendedRetry statement

  When you configure extended retry, mail messages that were undeliverable at the end of the long retry period are tried again for an extended period after the originating spool file is released. During the extended retry period, mail messages are saved in a z/OS UNIX file system directory. Because the originating spool file is released during extended retry processing, you can configure the ExtendedRetry statement to a longer time span than the time span that you configure for long retry. For example, the time span on the ExtendedRetry statement might be several days.

You can use these statements individually or at the same time to provide the best system environment for mail processing.

For example, the following statements set long retry to 1 hour and extended retry to 3 days:

```
RetryLimit
  {
  Count 6
  Interval 10
  }
ExtendedRetry
  {
  Age   3
  Interval 60
  MailDirectory /var/cssmtp/CSSMTP/mail/
  }
```

Tip: Because CSSMTP is a mail forwarder, the likelihood of receiving retry reply codes from a message transfer agent (MTA) is low. You are most likely to receive these errors when the MTA is also the home mail server for the mail recipients and the MTA reports errors that are related to the mailboxes of the recipients on the MTA. One possible error is that the mailbox is full.

Customizing the CSSMTP configuration file to handle undeliverable mail

CSSMTP provides configuration options to indicate how it handles undeliverable mail messages. You can configure CSSMTP to send individual undeliverable mail notifications or not to send them. You can also configure CSSMTP to create a report describing errors that were found while processing mail messages; this report can be created on the spool or sent to the configured mail administrators.

The following examples describe the configuration options for handling undeliverable mail. For more information about the BadSpoolDisp statement, the Undeliverable statement, the MailAdministrator statement, the ReportMailFrom statement, and the Report statement, see z/OS Communications Server: IP Configuration Reference.

- Example 1 — Configuring CSSMTP to not return an undeliverable mail notification to the originator and to hold the original spool file on the JES spool data set:

  ```
  BadSpoolDisp       Hold
  Report             Admin
  MailAdministrator  myuserId@ibm.us.com
  ReportMailFrom     userID@ibm.us.com
  Undeliverable
  {
    ReturnToMailFrom No
  }
  ```

  To configure the application to not send an undeliverable mail notification to the originator, you need to set the ReturnToMailFrom value to No. This example might be useful if the mail messages do not specify an originator, or if the spool file is for sending bulk mail that does not typically require a reply to the originator. This example has optionally configured that a report be sent to the specified mail administrator. If the Mail From of the error report is empty, it will use the mail address in ReportMailFrom if it has been configured.

  Tips:

  – The mail administrator can use the received report to determine which mail messages were undeliverable in the original spool file.

  – When the examination of this JES spool file is complete, the mail administrator should delete the JES spool file, which is now in a hold state.

  Results:

  – CSSMTP tries to send all mail messages, and if any of the mail is undeliverable, CSSMTP does not create and send an undeliverable mail notification.

  – Because the BadSpoolDisp statement is set to Hold, CSSMTP does not delete the JES spool file.

- Example 2 — Configure CSSMTP to send an undeliverable mail notification to the originator, and to hold the original spool file on the JES spool data set if necessary:

```
BadSpoolDisp       Hold
Report             None
Undeliverable
{
  ReturnToMailFrom     Yes
  DeadLetterAction     Store
  DeadLetterDirectory  /var/cssmtp/myDir/
}
```

This example is useful if the originator of the mail messages needs to be informed that the messages cannot be delivered. If you determine that it is not necessary to inform the originator of failed deliveries, set the ReturnToMailFrom value to No. No report is created in this example.

Tip: When spool files contain many mail messages, this configuration should be used with caution because a failure can cause many individual undeliverable mail notifications to be maintained in storage while trying to return them to the originator.

Results:

- CSSMTP builds an undeliverable mail notification and attempts to notify the originator of the mail message failure.
- Because the ReturnToMailFrom value is Yes, if the original spool file contains no errors other than undeliverable mail errors, CSSMTP always deletes the spool file even when the BadSpoolDisp value is set to Hold.
- If the original spool file contains both undeliverable mail errors and syntax errors, CSSMTP holds the spool file because the BadSpoolDisp value is set to Hold.
- If the undeliverable mail notification cannot be returned to the originator of the mail message, then this undeliverable mail notification becomes a dead letter. The action taken by CSSMTP is based on the value configured on the DeadLetterAction parameter. In this example, because the DeadLetterAction value is set to Store, CSSMTP stores the dead letters in the /var/cssmtp/myDir directory:

  ```
  /var/cssmtp/myDir/TESTMAIL.SYS00006.Sep302008.160454.541437.1U
  /var/cssmtp/myDir/TESTMAIL.SYS00006.Sep302008.160454.541999.1U
  ```

- Because None is specified on the Report statement, the log must be inspected for messages about any problems found in the JES spool file.

Steps for granting authority to start CSSMTP

Define explicit authority for all user IDs that you want to be able to start CSSMTP.

Procedure

Perform the following steps to grant authority to a user ID to start CSSMTP:

1. Ensure that the OPERCMDS class is active and RACLISTed, and that RACLIST processing is enabled.
2. Define the OPERCMDS class profile using a security product like RACF.
3. Grant CSSMTP access to the OPERCMDS class and then refresh the OPERCMDS class.

Results

Depending on how you start CSSMTP, some RACF profiles that restrict who can issue the START, CANCEL, MODIFY, or STOP operator commands are as follows:

```
MVS.START.STC.CSSMTP.*
MVS.START.STC.CSSMTP

MVS.CANCEL.STC.CSSMTP.*
MVS.CANCEL.STC.CSSMTP

MVS.MODIFY.STC.CSSMTP.*
MVS.MODIFY.STC.CSSMTP
```

```
MVS.STOP.STC.CSSMTP.*
MVS.STOP.STC.CSSMTP
```

For more information about protecting operator commands, see z/OS MVS Planning: Operations.

Security for CSSMTP

Consider the following additional security measures for CSSMTP:

- For spool files from NJE nodes, the user ID associated with the spool file must be defined by SAF. For more information about the protection of SYSOUT data sets, see the following topics:
  - Protecting Data Sets on Spools in z/OS Security Server RACF Security Administrator's Guide
  - Authorizing SYSOUT in z/OS Security Server RACF Security Administrator's Guide
  - Authorizing Network Jobs and SYSOUT (NJE) in z/OS Security Server RACF Security Administrator's Guide
  - Authorizing SYSOUT in z/OS JES2 Initialization and Tuning Guide
  - Understanding default userids in z/OS JES2 Initialization and Tuning Guide
  - Using RACF to Provide Security in z/OS JES3 Initialization and Tuning Guide
  - Understanding Default User IDs in z/OS JES3 Initialization and Tuning Guide
- If your installation protects access to the JESSPOOL class of resources, provide ALTER access to the CSSMTP user ID so that it can read and delete spool files. An example JESSPOOL definition follows:

```
//CSSMTP   EXEC PGM=IKJEFT01
//SYSTSPRT DD  SYSOUT=*
//* The PERMIT for CLASS(JESSPOOL) is needed only if it has already
//*   been activated.
//SYSTSIN  DD  *
    SETROPTS CLASSACT(STARTED)
    SETROPTS RACLIST(STARTED)
    SETROPTS GENERIC(STARTED)
    ADDUSER  CSSMTP  DFLTGRP(OMVSGRP) OMVS(UID(nn) HOME('/'))           -
             NOPASSWORD NAME('Simple Mail Transfer') OWNER(OMVSGRP)
    RDEFINE  STARTED OWNER(SYS1) CSSMTP.* STDATA(USER(CSSMTP))
    RDEFINE  JESSPOOL  localnodeid.** UACC(READ)
    PERMIT   localnodeid.**                                            -
             CLASS(JESSPOOL) ID(CSSMTP) ACCESS(ALTER)
    SETROPTS GENERIC(JESSPOOL) REFRESH
    SETROPTS RACLIST(STARTED) REFRESH
    SETROPTS GENERIC(STARTED) REFRESH
```

- When CSSMTP is defined with a nonzero UID value, Delete Operator Message (DOM) messages are prefixed with message BPXM023I. To remove the prefix, you must authorize the CSSMTP procedure user ID to use the UNIX System Services console service. You can authorize CSSMTP to use the console service by entering the following commands:

```
RDEFINE FACILITY BPX.CONSOLE UACC(NONE)
PERMIT BPX.CONSOLE -
CLASS(FACILITY) ID(CSSMTP) ACCESS(READ)
SETROPTS RACLIST(FACILITY) REFRESH
```

- You can control whether CSSMTP reads and processes the spool files created by specific users by creating one or more resource profiles in the SERVAUTH class.

  The format of the SERVAUTH profile name is EZB.CSSMTP.*sysname*.*writername*.*originJESnode*, where *sysname* is the system name defined in the sysplex, *writername* is the CSSMTP configured external writer name, and *originJESnode* is the JES node that originated the spool file. If this profile is created with UACC(NONE), then only user IDs permitted to the resource are able to have spool files processed by CSSMTP.

  For examples of the resource profile definitions, see the EZARACF sample in data set SEZAINST. For information about configuring the external writer name using the ExtWrtName statement, see z/OS Communications Server: IP Configuration Reference.

  Tips:

– You can specify a wildcard on segments of the profile name, as shown in the following example:

```
SETR GENERIC(SERVAUTH) GENCMDS(SERVAUTH)
SETR CLASSACT(SERVAUTH)
RDEFINE  SERVAUTH EZB.CSSMTP.sysname.writername.originJESnode          -
    UACC(NONE)
 PERMIT EZB.CSSMTP.sysname.writername.originJESnode                    -
    CLASS(SERVAUTH) ID(userid) ACCESS(READ)

 PERMIT EZB.CSSMTP.sysname.writername.*                                -
    CLASS(SERVAUTH) ID(userid) ACCESS(READ)

 SETROPTS RACLIST(SERVAUTH) REFRESH
```

– You can define multiple profiles. For example, if this CSSMTP instance processes spool files from local jobs and from remote systems, you can define a profile for the local system and for each JES node that originates spool files. The spool file is matched to the most specific RACF profile, and then the user ID associated with the spool file is validated against that profile.

Result: If the profile is active with UACC(NONE), then when the spool file is received on the JES spool data set, CSSMTP checks to determine whether a profile is defined for the originating node or for any originating node (*), and then checks whether the user ID associated with the job that created the spool file is permitted to the profile:

– If the user ID is permitted, the spool file is processed by CSSMTP.

– If the user ID is not permitted, the spool file is considered to be a bad spool file and is subject to the action specified by the BadSpoolDisp statement. For more information about the BadSpoolDisp statement, see z/OS Communications Server: IP Configuration Reference.

If the profile is not defined, then the spool file is processed as if the user ID is permitted.

- You can write a CSSMTP user exit or use the existing SMTPD user exit to inspect mail before it is sent to a target server.

  For example, you can create an exit to check the MAIL FROM: string on outbound mail. The mail originator, recipients, and other information can be inspected by the exit, and the entire JES job or a single mail message can be discarded.

  You can control the user exit that is used with the UserExit statement:

  – If you want to use a user exit that you have written for CSSMTP, then specify Version3 on the UserExit statement. You need to use this version if you want to use both RFC 821 and RFC 2821 commands to read and process from the JES spool data set. For more information about the CSSMTP exit, see z/OS Communications Server: IP Configuration Reference.

  – If you want to continue using your SMTPD user exit and use only RFC821 commands, then specify Version2 on the UserExit statement.

  For more information about the UserExit statement, see z/OS Communications Server: IP Configuration Reference.

- You can enable the SMTP server and client to use Transport Layer Security (TLS) to provide private, authenticated communication over the Internet using RFC 3207, *SMTP Service Extension for Secure SMTP over Transport Layer Security*. For more information, see “Steps for using Transport Layer Security for CSSMTP” on page 1293.

Steps for using Transport Layer Security for CSSMTP

You can enable the SMTP server and client to use Transport Layer Security (TLS) to provide private, authenticated communication over the Internet.

Procedure

Perform the following steps to use Transport Layer Security (TLS) for CSSMTP:

1. Set up secure mail using the YES option on the Secure parameter of the TargetServer statement, the STARTTLS command in the JES batch job, or both.

Table 72 on page 1294 shows whether TLS is required between CSSMTP and a target server based on various JES batch job and CSSMTP configuration combinations.

Table 72. JES batch job and CSSMTP configuration combinations for secure mail

STARTTLS command?	Secure parameter value	TLS required?
Yes	YES	Yes
No	YES	Yes
Yes	NO	Yes
No	NO	No

If secure communication is required according to Table 72 on page 1294 and no available target servers support TLS (as indicated by server capabilities in response to the EHLO command), the mail message fails and is not delivered.

For information about the TargetServer statement, see z/OS Communications Server: IP Configuration Reference. For information about the STARTTLS command, see z/OS Communications Server: IP User's Guide and Commands.

2. See the following simple example to get started with TLS.

   For more information about TLS, see Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121.

   In this example, assume the following characteristics:

   - The mail contains sensitive data, and you want CSSMTP to communicate with only TLS protocols.
   - CSSMTP is using port 25 to communicate with a target server on another platform.
   - There is only one TCP/IP stack over which mail is delivered, referred to as the client stack.

   To set up TLS for this sample environment, take the following actions:

   a. Create the key ring.

      The client key ring needs the root certification used to sign the server certificates. For a TLS/SSL primer and some step-by-step examples, see Appendix B, "TLS/SSL security," on page 1329. For more information about managing key rings and certificates with RACF and the RACDCERT command, see z/OS Security Server RACF Security Administrator's Guide. For more information about managing key rings and certificates with gskkyman, see z/OS Cryptographic Services System SSL Programming.

   b. Configure CSSMTP to require secure communication. Configure the TargetServer statement with the Secure parameter set to YES, which specifies that TLS protocols are always required. For information about the TargetServer statement, see z/OS Communications Server: IP Configuration Reference.

   c. Configure the client system to use TLS with AT-TLS policies as follows:

      i) Specify TTLS on the TCPCONFIG statement in the TCP/IP profile for the client stack. For information about the TCPCONFIG statement, see z/OS Communications Server: IP Configuration Reference.

      ii) Block the ability of applications to open a socket before AT-TLS policy is loaded into the TCP/IP stack by setting up EZB.INITSTACK.*sysname.tcpname* for the client stack.

      iii) Create a main Policy Agent configuration file containing a TcpImage statement for the client stack, and create a TcpImage policy file for the client stack. For more information about AT-TLS policy statements, see z/OS Communications Server: IP Configuration Reference.

      iv) Add a TTLSConfig statement to each TcpImage policy file to identify the TTLSConfig policy file location:

```
TTLSConfig  clientPath
```

v) Add the AT-TLS policy statements to the *clientPath* file:

```
TTLSRule                          CSSMTPRule
{
   RemotePortRange                25
   Direction                      Outbound
   TTLSGroupActionRef             CSSMTPGroup
   TTLSEnvironmentActionRef       CSSMTPEnvironment
}
TTLSGroupAction                   CSSMTPGroup
{
   TTLSEnabled                    On
}
TTLSEnvironmentAction             CSSMTPEnvironment
{
   HandshakeRole Client
   TTLSKeyRingParms
   {
      Keyring                     CSSMTP/client_key_ring
   }
   TTLSEnvironmentAdvancedParms
   {
      ApplicationControlled       On
   }
}
```

Tip: You can use the IBM Configuration Assistant for z/OS Communications Server to generate the AT-TLS Policy Agent files. For information about the IBM Configuration Assistant for z/OS Communications Server, see "Option 1: Use the IBM Configuration Assistant for z/OS Communications Server" on page 1122.

3. If the server requires an EHLO command to be sent after a successful TLS negotiation, configure TLSEhlo Yes on the Options statement in the CSSMTP configuration. For more information, see Options statement in z/OS Communications Server: IP Configuration Reference.

Results

You know you are done when CSSMTP can successfully deliver mail to a target server using secure connections. If SECURE YES is configured and CSSMTP is able to successfully negotiate and establish a TLS session, the following message is displayed:

```
EZD1821I  csproc ABLE TO USE TARGET SERVER  ipAddress
```

Restriction: To use the STARTTLS command with a target server, the target server must have a certificate that can be validated by the AT-TLS component of z/OS Communications Server as configured by Policy Agent. This certificate can be a self-signed certificate or a certificate that can be validated by a known certificate authority. If the certificate of the server cannot be validated, secure communication with the server fails and mail that requires security cannot be delivered to that server.

Steps for configuring SMF records for CSSMTP (optional)

You can use SMF to record CSSMTP events.

Before you begin

CSSMTP can write System Management Facilities (SMF) records for the following cases:

- When the configuration has been initialized or modified by a MODIFY REFRESH command, a MODIFY LOG,LEVEL command, or a MODIFY USEREXIT command. A CSSMTP application configuration record (CONFIG subtype 48) is written.
- At the end of a connection with a target server, a CSSMTP application connection record (CONNECT subtype 49) is written.
- At the end of the processing for each mail message, a CSSMTP application mail message record (MAIL subtype 50) is written.
- At the end of the processing of a spool file when all the mail messages have been completed, a CSSMTP application spool file record (SPOOL subtype 51) is written.

- When statistical information about the CSSMTP processing at the SMF intervals is required, a CSSMTP application statistical record (STATS subtype 52) is written.

Procedure

Perform the following steps to write SMF records to both the MVS SMF data sets and the real-time SMF NMI. Writing SMF records to the MVS SMF data sets and writing SMF records to the real-time SMF NMI are two independent functions and are controlled by different configuration parameters.

1. Add a SMF119 statement to the CSSMTP configuration file.

   This configuration parameter controls which SMF record types are written to the MVS SMF data sets. For example:

   ```
   SMF119
       {
       CONFIG YES
       SPOOL YES
       MAIL NO
       CONNECT YES
       STATS YES
       }
   ```

   In this example, the CONFIG, SPOOL, CONNECT and STATS records will be written to MVS SMF data sets and not the MAIL record.
2. Restart the CSSMTP task or issue a MODIFY REFRESH command to dynamically update CSSMTP with the new SMF119 statement values.
3. Update the TCP/IP profile NETMONITOR statement to forward all of the CSSMTP SMF records that are associated with the CONFIG, SPOOL, CONNECT, and STATS types to an application using the real-time SMF NMI, SYSTCPSM.

   These record types are both connection oriented and non-connection oriented (see “5” on page 1296). For more information about SYSTCPSM, see z/OS Communications Server: IP Programmer's Guide and Reference.

   ```
   NETMONITOR SMFSERVICE CSSMTP
   ```

   For the descriptions of the SMF records see z/OS Communications Server: IP Programmer's Guide and Reference.
4. Update the TCPIP profile NETMONITOR statement to forward all the CSSMTP SMF records associated with the MAIL type to an application using the real-time SMF NMI, SYSTCPSM.

   ```
   NETMONITOR SMFSERVICE CSMail
   ```

   For more information about SYSTCPSM, see z/OS Communications Server: IP Programmer's Guide and Reference.
5. Special considerations for CSSMTP application and NETMONITOR in a CINET environment:
   - The CSSMTP application can be started with the -p parameter to set stack affinity. All the SMF records are written to the real time SMF NMI associated with the stack whose name was specified on the -p parameter. This forces connection oriented SMF records (CONNECT) and non-connection oriented SMF records (CONFIG, SPOOL, MAIL and STATS) to go to the same stack. The CSSMTP and CSMAIL NETMONITOR parameters should be specified in the profile data sets of the stack name to match the name specified on the -p parameter.
   - If the CSSMTP application is started without the -p parameter and multiple stacks are active, then a network management application can not determine the stack that records will be written to. The CSSMTP and CSMAIL NETMONITOR parameters should be specified in the profile data sets of all stacks, so that network management applications can obtain all the records. The network management application will not get redundant records. In this case CSSMTP writes the records to the first stack with an active SMFService application and with the appropriate SMFSERVICE CSMAIL or CSSMTP parameter set.

Monitoring CSSMTP

You can use the following MODIFY commands to monitor CSSMTP:

- MODIFY DISPLAY,CONFig

  Displays the current configuration values.

- MODIFY DISPLAY,IPList

  Displays the current set of target servers to which mail can be sent that are configured or resolved from the TargetServer statement.

- MODIFY DISPLAY,SPoolstatus

  Helps you determine how mail is being processed off the JES spool file.

- MODIFY DISPLAY,TARgets

  Helps you monitor the mail that is being sent to target servers to determine whether mail is being delivered, is being placed in undeliverable states, or is being tried again through long retry queues. This command also helps you determine the list of target servers, their attributes, and the number of mail messages that a target server is accepting.

For information about the MODIFY command for CSSMTP, see z/OS Communications Server: IP System Administrator's Commands.

sendmail to CSSMTP bridge

The sendmail to CSSMTP bridge (sendmail bridge) allows the user to send emails by using the facilities of z/OS UNIX. The sendmail bridge command parses input command switches, reads the mail message from the UNIX System Services file, and processes the mail message. The input mail message is updated by adding SMTP commands and SMTP headers, if no headers are specified in the input mail message. The updated mail message is transmitted to the JES spool data set that the Communications Server SMTP (CSSMTP) application can process. CSSMTP must be configured and running.

Restriction: The sendmail bridge is only a mail client function.

This topic provides information about how to configure the sendmail bridge.

For more configuration information, see Sendmail to CSSMTP bridge in z/OS Communications Server: IP Configuration Reference and for sendmail bridge command information, see z/OS Communications Server: IP User's Guide and Commands.

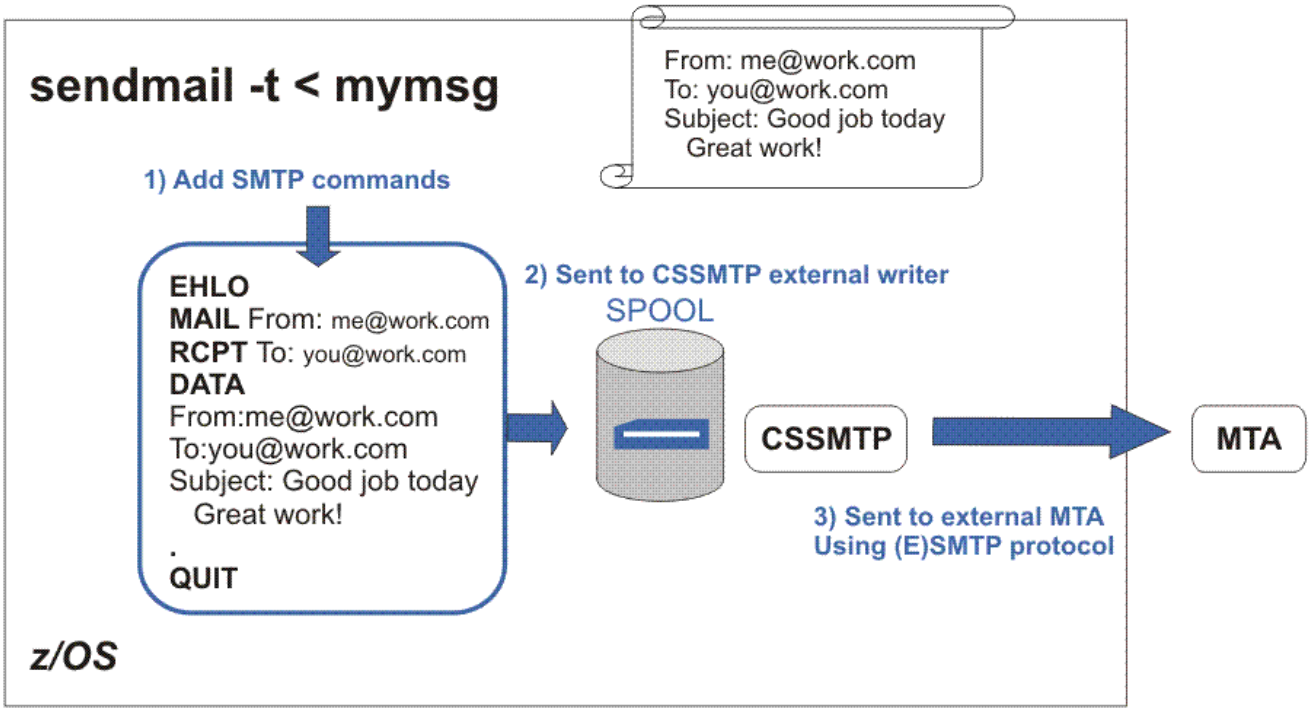


Figure 157. sendmail to CSSMTP bridge

Setting up the sendmail bridge

To set up the sendmail bridge to process and forward your mail, you must take the following actions:

1. Create and customize the configuration file.

   See "Steps for creating and customizing the configuration file" on page 1298.
2. Configure and start CSSMTP.
   - Issue F *cssmtpproc*,display,config to check if CSSMTP is running
   - If you have not configured and started CSSMTP , see Configuring the CSSMTP application.
3. Consider security for the sendmail bridge.

   See "Security considerations for the sendmail bridge" on page 1299.

Steps for creating and customizing the configuration file

You can take the following actions to create and customize the sendmail bridge configuration file:

- Copy the sample file ezatmail.cf from /usr/lpp/tcpip/samples/ to /etc/mail/ezatmail.cf.

  Guideline: If you used z/OS sendmail in previous releases, you can use your existing configuration files with the sendmail bridge: /etc/mail/submit.cf or /etc/mail/sendmail.cf. However, because many statements in those 2 files are not supported for the sendmail bridge, it is suggested that you use ezatmail.cf instead.

  You can use the sendmail bridge command line switch -C to define the configuration file. For the search order of the sendmail bridge configuration file, see z/OS Communications Server: IP Configuration Reference.
- Customize the configuration file:

- If you plan to use an alias name for your mail recipients, you might need to provide an O statement with options AliasFile and MaxAliasRecursion.
- If you plan to limit the number of recipients in one mail message, you might need to provide an O statement with option MaxRecipientsPerMessage.
- If you plan to use a different CSSMTP external writer name from the default name "CSSMTP", you can configure a W statement with the CSSMTP external writer name that matches the name of your CSSMTP application. You can also use the environment variable EZATMAIL_CSSMTP_EXTWRTNAME to define your CSSMTP external writer name. For the search order of the sendmail bridge CSSMTP external writer name, see z/OS Communications Server: IP Configuration Reference.
- You can customize options for the sendmail bridge command. For additional information, see z/OS Communications Server: IP User's Guide and Commands.

Security considerations for the sendmail bridge

- With the sendmail bridge command, the updated input mail message is transmitted to the JES spool data set that the Communications Server SMTP (CSSMTP) application processes. For more information about the protection of SYSOUT data sets, see Security for CSSMTP. You can request Transport Layer Security (TLS) for messages sent by the sendmail bridge command through the configuration statement D{tls_version}. If D{tls_version} is specified, a STARTTLS SMTP command is added. When CSSMTP processes the JES spool data set, CSSMTP ensures a secured connection between CSSMTP and its target server is used to forward the mail message. The value that is defined by D{tls_version} is ignored. The secured connection is setup between CSSMTP and the target mail server based on configured AT-TLS policy.
- It is not required that D{tls_version} is set in the configuration file to ensure a secured connection is used. The user can ensure that a secured connection is used between CSSMTP and its target server by setting the Secure parameter to Yes on the CSSMTP TargetServer configuration statement. See “Steps for using Transport Layer Security for CSSMTP” on page 1293.
- If you plan to invoke the sendmail bridge command by submitting a batch job that uses BPXBATCH, the sendmail bridge assumes user IDs are used when running the command, and these user IDs must be defined to execute sendmail correctly. The commands to define the sendmail user IDs are defined in SEZAINST(EZARACF). The commands are:

```
RDEFINE STARTED  SENDMAIL.* STDATA(USER(SENDMAIL))
SETROPTS RACLIST(STARTED) REFRESH
SETROPTS GENERIC(STARTED) REFRESH
```

Configuring popper

POP3 uses port 110. You can define additional ports if there is a need for additional command-line options for popper. For information on the options that might be suitable for your site, see the z/OS Communications Server: IP User's Guide and Commands.

z/OS UNIX popper will most likely be used by those whose local mailer requires a POP3 server. Typically their administrator will provide them with the address or name of the z/OS running the POP3 server, with instructions on where this information should be used.

The authentication method used in this implementation of popper is user ID and password.

The popper implementation can be divided into four steps:

- Update the /etc/services file.
- Update the /etc/inetd.conf file.
- Create the directory for the temporary maildrop file.
- Start inetd.

Update the /etc/services file

You need a port for the POP3 server defined in /etc/services. Add the following line to your /etc/services file. Note that the well-known port 110 is being used in this example:

```
pop3               110/tcp popper
```

Update the /etc/inetd.conf file

Because popper is invoked by INETD, add the following information to your /etc/inetd.conf file, where -d indicates to run popper in debugging mode:

```
pop3       stream tcp nowait bpxroot     /usr/sbin/popper popper -d
```

The debugging option is used to confirm correct installation, as shown in "Correct connection" on page 1300.

Create the directory for the temporary maildrop file

When popper is invoked through a Mail User Agent (MUA) client request, such as GET NEW MESSAGES, popper starts to read the user's mail file system (/usr/mail/username) where the data has been stored, and puts this data (if there are messages) in a temporary maildrop file /usr/mail/popper/.username.pop. The contents of this file are transmitted to the remote client using the existing POP3 TCP connection.

Because the directory does not exist, create it as follows:

```
/usr/mail/popper/
chmod 777 /usr/mail/popper
```

The popper uses this directory to create and fill the maildrop file. If this directory is not specified or if permissions are incorrect, you will get an error message similar to the following message:

```
EZZ7605I:Unable to open temporary maildrop '/usr/mail/popper/.username.pop'
```

Also, inside the POP3 session, you will get an error message similar to the following message:

```
-ERR System error, can't open temporary file, do you own it?
```

Restrictions:

- If the user maildrop file is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the popper.
- If the user maildrop file is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which the user maildrop file is stored cannot have write access to the directory.

Start inetd

After updating the inetd.conf file, start INETD. For more information on inetd, see Appendix A, "Setting up the InetD configuration file," on page 1327. Also see z/OS Communications Server: IP Configuration Reference.

Correct connection

The following example shows a working connection to the popper server:

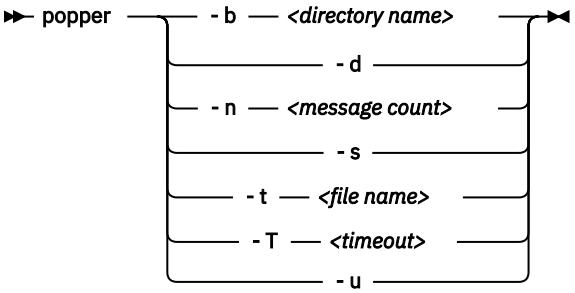
```
Debugging turned on
(v2.3)Servicing request from "hostname.domainname"at xxx.xxx.xxx.xxx
1
+OK QPOP (version 2.3)at hostname.domainname starting.
Sending line "+OK QPOP (version 2.3)at hostname.domainname starting."
Received:"USER username"
+OK Password required for username.
```

```
Sending line "+OK Password required for username."
Received:"pass xxxxxxxxx"
Creating temporary maildrop '/usr/mail/popper/.username.pop'
uid =4029,gid =1
Checking for old .username.pop file
Old .username.pop file not found,errno (0)
Msg 1 being added to list
Msg 1 uidl c313e341d7a5167b833153eb4bbfea25 at offset 0 is 700 octets long and h
Msg 2 being added to list
Msg 2 uidl c50b65c95f934fb6c22dc23573be88a1 at offset 748 is 2072 octets long an
Msg 3 being added to list
Msg 3 uidl 91c0636d1513127489f49995e6d8f1e5 at offset 2842 is 3998 octets long a
Msg 4 uidl 61021c4ea6471f83f518d8c64d8c1740 at offset 6772 is 453814 octets long
Msg 5 being added to list
Msg 5 uidl da701c81e2b2df3a60121a5ca1cdd76b at offset 454502 is 928 octets long
Msg 6 being added to list
Msg 1 uidl c313e341d7a5167b833153eb4bbfea25 at offset 0 is 700 octets long and h
Msg 2 uidl c50b65c95f934fb6c22dc23573be88a1 at offset 748 is 2072 octets long an
Msg 3 uidl 91c0636d1513127489f49995e6d8f1e5 at offset 2842 is 3998 octets long a
Msg 4 uidl 61021c4ea6471f83f518d8c64d8c1740 at offset 6772 is 453814 octets long
Msg 5 uidl da701c81e2b2df3a60121a5ca1cdd76b at offset 454502 is 928 octets long
Msg 6 uidl 0a48082f727723c8f47b306e26b49652 at offset 455469 is 691 octets long
+OK username has 6 messages (462203 octets).
Sending line "+OK username has 6 messages (462203 octets)."
Received:"STAT"
6 message(s)(462203 octets).
+OK 6 462203
Sending line "+OK 6 462203"
Received:"LIST"
+OK 6 messages (462203 octets)
Received:"UIDL"
+OK uidl command accepted.
Sending line "+OK uidl command accepted."
Sending line "1 c313e341d7a5167b833153eb4bbfea25"
Sending line "2 c50b65c95f934fb6c22dc23573be88a1"
Sending line "3 91c0636d1513127489f49995e6d8f1e5"
Sending line "4 61021c4ea6471f83f518d8c64d8c1740"
Sending line "5 da701c81e2b2df3a60121a5ca1cdd76b"
Sending line "6 0a48082f727723c8f47b306e26b49652"
Received:"QUIT"
Performing maildrop update...
Checking to see if all messages were deleted
Opening mail drop "/usr/mail/username"
Creating new maildrop "/usr/mail/username"from "/usr/mail/popper/.username.pop"
Copying message 1.
Copying message 2.
Copying message 3.
Copying message 4.
Copying message 5.
Copying message 6.
+OK Pop server at hostname.domainname signing off.
Sending line "+OK Pop server at hostname.domainname signing off."
(v2.3)Ending request from "username"at (hostname.domainname)xxx.xxx.xxx.xxx
```

Popper command - administering received mail

If the receiver MUA does not have direct access to the mail spool file, use Popper to access the mail spool on the local host. z/OS Popper is used when a POP3 server is needed.

Syntax



Parameters

The following command-line options can be used when you are invoking Popper.

-b *<directory name>*
Specifies the name of the directory in which bulletins are found. If not specified, `/usr/mail/bulletins` is used as the default.

Restrictions:

- If the bulletins file is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to Popper.
- If the bulletins file is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which the bulletins file is stored cannot have write access to the directory.

-d
Requests more debugging messages be turned on.

-n *<message count>*
Specifies the number of old bulletins to be delivered to new users. If not specified, no bulletins are delivered.

-s
Requests statistics logging be turned on.

-t *<file name>*
Specifies a trace file for all message logging. If not specified, messages are logged via the syslog facility.

-T *<timeout>*
Specifies the time, in seconds, before an idle POP3 connection is terminated. RFC 1939 specifies a minimum timeout of 600 seconds, but in practice such a long timeout does not work well. (When a connection gets aborted, the user is locked out of their mailbox for the timeout period.) If not specified, 120 seconds is used as the default timeout period.

-u
Requests that the user mailbox be updated on abort. RFC 1939 specifies that mailboxes should not be updated (that is, no messages should be deleted) if a connection is aborted abnormally. This option forces an update to occur despite the aborted connection. If not specified, no update occurs on aborted connections.

Chapter 27. TIMED daemon

TIMED is a daemon that is used to provide the time in response to UDP requests. TIMED gives the time in seconds since midnight January 1, 1900. You can start TIMED from the z/OS shell or as a started procedure. TCP/IP must be started prior to starting TIMED.

Note: TIMED is different from the TIME daemon available as an internal daemon of INETD. INETD cannot be used to start and perform as a listener for TIMED.

Starting TIMED from the z/OS shell

TIMED is installed in the /usr/lpp/tcpip/sbin/ directory.

To start the TIMED server from the command line, issue the **timed** command.

```
timed [-l] [-p port]
```

The following parameters are used for the **timed** command:

-l
Logs all the incoming requests and responses to the system log. Logged information includes the IP address of the requestor.

-p port
The TIMED server usually receives requests on well-known port 37. TIMED uses UDP only. You can specify the port in which requests are to be received.

Starting TIMED as a procedure

The following sample shows how to start TIMED as a procedure.

```
//TIMED    PROC
//*
//*   5694-A01 (C) Copyright IBM Corp. 1997, 2002
//*   Licensed Materials - Property of IBM
//*   This product contains "Restricted Materials of IBM"
//*   All rights reserved.
//*   US Government Users Restricted Rights -
//*   Use, duplication or disclosure restricted by
//*   GSA ADP Schedule Contract with IBM Corp.
//*   See IBM Copyright Instructions.
//*
//* Function: Time server start procedure
//* SMP/E distribution name: EZATTMDP
//*
//TIMED    EXEC PGM=TIMED,REGION=0K,TIME=NOLIMIT,
//         PARM='POSIX(ON),ALL31(ON),TRAP(OFF)/'
//*STEPLIB DD DISP=SHR,DSN=TCP.SEZALOAD,
//*           VOL=SER=,UNIT=
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
//SYSOUT   DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
//SYSABEND DD SYSOUT=*
//         PEND
```

© Copyright IBM Corp. 2000, 2022

Chapter 28. SNTP daemon

The SNTP daemon (SNTPD) is a TCP/IP daemon that is used to synchronize time between a client and a server. Simple Network Time Protocol (SNTP) is a protocol for synchronizing clocks across a WAN or LAN through a specific formatted message. An External Time Reference (ETR), named stratum 0, is chosen as the highest timer reference used for synchronization. A stratum 1 server is attached to and receives the time from the stratum 0 timer. For example, the z/OS sysplex timer could be a stratum 0 timer, and z/OS Communications Server would be a stratum 1 server. A client attached to the stratum 1 server can also be a stratum 2 server, receiving the time from the stratum 1 server, and so on. SNTP uses UDP packets for data transfer with the well-known port number 123. RFC 2030 (Mills 1996) describes SNTP. You can start SNTPD from the z/OS shell or as a started procedure. TCP/IP must be started prior to starting SNTPD.

Restrictions:

- The SNTP server only processes requests with a Mode value of 1 through 4. The SNTP server discards requests whose Mode value is 0, or is greater than 4.
- As per RFC 2030, the SNTP server discards requests whose length is less than the minimum required length of 48 bytes.

SNTPD is an SNTP server. If the server is configured for unicast mode only (that is, neither the multicast mode or the broadcast mode start parameters are specified), then the server just waits for requests from clients and then responds. If the server is also configured for multicast or broadcast mode, it sends unsolicited messages at the intervals that are specified on the multicast mode or the broadcast mode start parameters.

The /etc/sntpd.pid is a temporary SNTP daemon pid file that the SNTP daemon creates. This file contains the process ID of the current invocation of the SNTP daemon.

Restrictions:

- If /etc/sntpd.pid is a symbolic link, it must have an owning UID or GID that matches the EUID or EGID that is assigned to the SNTP daemon.
- If /etc/sntpd.pid is a hard link or the target of a hard link, users that are outside the owner or group of the directory in which /etc/sntpd.pid is stored cannot have write access to the directory. Additionally, write access to /etc/sntpd.pid must be limited to the owning UID or group, for example, --w--w---- permissions.

Steps for starting SNTPD from the z/OS UNIX shell

You can start SNTPD from the z/OS UNIX shell.

Before you begin

Ensure the existence of the following files. The files used by z/OS UNIX SNTPD and their locations in the z/OS UNIX file system are as follows:

/etc/services
The ports for each application are defined here.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file.

/usr/sbin/sntpd
This is a symbolic link to /usr/lpp/tcpip/sbin/sntpd, which is a sticky-bit file. The SNTPD member of SEZALOAD contains the executable code for the SNTP server.

/usr/lib/nls/msg/C/sntpdmsg.cat
The message catalog used by the z/OS UNIX SNTPD server.

When restricting low port usage, the port used by SNTPD (default value of 123) should either:

© Copyright IBM Corp. 2000, 2022

- Be reserved for the name of the SNTPD start procedure
- Use the SAF parameter on the PORT statement to restrict access to the SNTPD port

There is no configuration file specifically for SNTPD.

Procedure

Type sntpd & on the command line.

This command starts SNTPD and sends it to the background. The following optional parameters are used for the sntpd command:

-d
Enables debugging. Debug messages go to syslog daemon.

-df *pathname*
Enable debugging. Debug messages go to the specified file location.

-pf *pathname*
Path for pid file. For example:

```
-pf /var
```

-m *nnnnn*
Acts in multicast mode. Sends multicast updates (TTL=1) on all interfaces every *nnnnn* seconds. Listens to multicast requests and responds with unicast replies. The valid range for `-m` is 1 to 16284.

-b *nnnnn*
Acts in broadcast mode. This parameter sends local broadcasts on all interfaces every *nnnnn* seconds. It also specifies to listen to broadcast requests and respond with unicast replies. The valid range for `-b` is 1 to 16284.

-s *nn*
Use *nn* as the stratum level in all replies sent by the server. The valid range for *nn* is 1 to 15. If `-s` is not specified or a value is specified that is not valid, the default stratum level of 1 is used. The stratum level indicates the relative accuracy of the local clock compared to the clocks of other SNTP servers in the network. The value 1 is most accurate and 15 is least accurate.

-?
Display the syntax of the command usage and options.

Results

You know that SNTPD has started when the following message appears on the MVS console:

```
EZZ9600I SNTP SERVER READY.
```

Steps for starting SNTPD as a procedure

You can start SNTPD as an MVS started procedure.

Before you begin

Obtain a copy of this sample procedure from SEZAINST and store it in one of your PROCLIB concatenation data sets.

Procedure

Invoke the procedure using the system operator start command.

The following sample, SEZAINST(SNTPD), shows how to start SNTPD as a procedure:

```
//SNTPD    PROC
//*
```

```
//* Sample procedure for the Simple Network Time Protocol (SNTP)
//*
//* z/OS Communications Server Version 1 Release 13
//* SMP/E Distribution Name: SEZAINST(EZASNPRO)
//*
//* Copyright:    Licensed Materials - Property of IBM
//*               5650-ZOS
//*               Copyright IBM Corp. 2002, 2015
//*
//* Status:       CSV2R2
//*
//SNTPD    EXEC PGM=SNTPD,REGION=4096K,TIME=NOLIMIT,
//         PARM='/ -d'
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//SYSIN    DD DUMMY
//SYSERR   DD SYSOUT=*
//SYSOUT   DD SYSOUT=*,DCB=(RECFM=F,LRECL=132,BLKSIZE=132)
//CEEDUMP  DD SYSOUT=*
//SYSABEND DD SYSOUT=*
//*
```

Results

You know that SNTPD has started when the following message appears on the console:

```
EZZ9600I SNTP SERVER READY.
```

Stack affinity

If you are running in a multiple stack environment and want SNTPD to use only a single stack, the environment variable _BPXK_SETIBMOPT_TRANSPORT can be used.

Chapter 29. Remote Execution

This topic describes how to configure and operate both the Remote Execution server and the UNIX Remote Execution server. z/OS Communications Server supports remote execution daemons in both the UNIX and TSO environments.

To execute commands under the UNIX shell, use the RSH command. To execute commands under TSO, use the REXEC command. These requests are serviced by the UNIX daemons, orshd and orexecd, and the TSO RXSERVE daemon.

The differences between the UNIX daemons and the TSO RXSERVE daemon are as follows:

- The UNIX daemons are initiated through the INETD server and can be configured to support a port other than their well-known port.
- The TSO daemon must be active and will service REXEC and RSH requests only on their well-known ports.

Only the UNIX daemons or the TSO daemon can be active at any one time.

UNIX REXEC

The UNIX Remote Execution Protocol Daemon (REXECD) is the server for the REXEC routine. REXECD allows execution of z/OS UNIX commands with authentication based on user names and passwords.

The Remote Shell Server (RSHD) is the server for the remote shell (RSH) client. The server provides remote execution facilities with authentication based on privileged port numbers, user IDs, and passwords.

See “Configuring the z/OS UNIX Remote Execution servers” on page 1313 for more information about configuring this server.

TSO REXEC

The TSO Remote Execution server allows execution of a TSO command that has been received at a remote host. This server runs the Remote EXEcution Command Daemon (REXECD) which supports both the Remote Execution (REXEC) and Remote Shell (RSH) protocols.

The TSO Remote Execution server has affinity for a specific transport in a CINET environment. Configure and execute a unique instance of the server for each TCP/IP stack requiring TSO remote execution services.

Configuring the TSO Remote Execution server

Procedure

Perform the following steps to configure the TSO Remote Execution server:

1. Update AUTOLOG and PORT statements in the PROFILE.TCPIP data set.
2. Determine whether the Remote Execution client will send a Remote Execution (REXEC) command or Remote Shell (RSH) command.
3. Permit remote users to access MVS resources. (Required only if the client is not sending a password.)
4. Update the Remote Execution cataloged procedure.
5. Create a user exit routine (optional).
6. Permit access to JESSPOOL files.

© Copyright IBM Corp. 2000, 2022

Step 1: Configuring PROFILE.TCPIP for TSO Remote Execution server

If you want the Remote Execution server to start automatically when the TCPIP address space is started, include the name of the member containing the RXSERVE cataloged procedure in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP data set.

```
AUTOLOG
  RXSERVE
ENDAUTOLOG
```

To ensure that port 512 is reserved for the Remote Execution protocol and port 514 for the Remote Shell protocol, add the name of the member containing the Remote Execution cataloged procedure to the PORT statement in *hlq*.PROFILE.TCPIP:

```
PORT
  512 TCP RXSERVE
  514 TCP RXSERVE
```

See z/OS Communications Server: IP Configuration Reference for more information about the AUTOLOG and PORT statements.

Step 2: Determine whether Remote Execution client will send REXEC or RSH commands

The Remote Execution client can send commands to the TSO Remote Execution server by the following methods:

1. Sending the Remote Execution (REXEC) command
2. Sending the Remote Shell (RSH) command with a user ID and password separated by a slash (/) character with the `-l` option on the RSH command
3. Sending the Remote Shell (RSH) command without a password

With methods "1" on page 1310 and "2" on page 1310, the TSO Remote Execution server executes the request and passes the password to MVS for verification. (REXEC commands require a password.) When these methods are used, skip Step 3.

With method "3" on page 1310, to enable an RSH client to send RSH commands to the TSO Remote Execution server without specifying a password, Step 3 is required.

Step 3: Permit remote users to access MVS resources (optional)

This step is necessary only if your installation allows users to issue remote execution commands without the requirement of specifying a password on the remote execution client.

Procedure

Use the following steps to ensure that the server can correctly access necessary MVS resources. You can use z/OS Security Server (RACF) or an equivalent security program.

1. Verify that your system has been configured for allowing surrogate job submission as described in z/OS Security Server RACF Security Administrator's Guide (SC28-1915) or by using an equivalent security program.
2. Authorize the TSO Remote Execution server to submit jobs for the MVS user ID specified with the `-l` option of the RSH command. This can be done with the RACF facility as described in z/OS Security Server RACF Security Administrator's Guide (SC28-1915), or by using an equivalent security program.
3. Define an *mvs_userid*.RHOSTS.DATA data set and authorize the TSO Remote Execution server user ID permission to read this data set.

   This can be done with the RACF facility as described in z/OS Security Server RACF Security Administrator's Guide (SC28-1915), or by using an equivalent security program.

   Note: This is the user ID used to start the RXSERVE address space.

This data set identifies the Remote Execution clients that can execute MVS commands remotely by sending an RSH command.

When a Remote Execution client sends an RSH request to the TSO Remote Execution server, the request includes the local user ID of the client user (*local_userid*) and, if the client user specified the `-l` option of the RSH command, the request also contains the user ID to use on the remote host (*mvs_userid*). If the client does not specify the `-l` option, the user ID to be used on the remote host is assumed to be the same as the *local_userid*.

When the TSO Remote Execution server receives an RSH command without a password, the server looks for a data set called *mvs_userid*.RHOSTS.DATA. The *mvs_userid*.RHOST.DATA data set contains one or more entries. Each entry consists of two parts, a fully qualified name of the client user's host and a *local_userid* associated with that host. The *local_userid* is case sensitive. If the data set exists, the server reads it and looks for an entry with a host name that matches the client user's host. If the user ID specified on this entry in the RHOSTS.DATA data set matches the *local_userid* passed on the RSH command, the RSH command continues processing. If the entry does not exist, the server responds to the client with message `EZA4386E Permission denied`.

Tip: If the client connected to this host through a link-local address, the client's host name generated by the resolver can be in the format *hostname%scope*. Adding *%scope* information to the appropriate RHOSTS.DATA client host definitions results in a more efficient search for a matching client host name. For details on the support for including scope information on configured host names, see z/OS Communications Server: IPv6 Network and Application Design Guide.

In the following example of an RHOSTS.DATA data set, the MVS client user `mvsuser` is allowed to issue the RSH command without a password from host `rs60007` with a local AIX user ID of `aixuser`.

Example of mvsuser.RHOSTS.DATA data set:

```
rs60007.itso.ral.ibm.com aixuser
```

Guideline: It is suggested not enabling the MVRSHD server. The MVRSHD server supports the RSH and REXEC protocols which transfer user ID and password information in the clear. There is also the potential of weak authentication for RSH clients using RHOSTS.DATA datasets. This authentication method allows remote command execution without requiring the RSH client to supply a password. IBM Health Checker for z/OS can be used to check whether the MVRSHD server is active and detect an RSH client attempting to use an RHOSTS.DATA dataset for authentication. For more details about IBM Health Checker, see IBM Health Checker for z/OS: User's Guide.

4. Users may be authenticated using Kerberos or GSS. If the username in the Kerberos or GSS credentials matches the local user ID (*local_userid*) of the client supplied by the RSH client, then no password is required.

Step 4: Update the TSO Remote Execution cataloged procedure

Update the TSO Remote Execution cataloged procedure by copying the sample provided in SEZAINST(RXPROC) to your system or recognized PROCLIB and modifying it to suit your local conditions. Specify the TSO Remote Execution server parameters and modify the JCL as required for your installation.

Rule: The PREFIX specification must be unique for every instance of RXSERVE that runs in a JESPLEX. The JESPLEX is a system that shares the same JES spool.

Tip: You can update the TSO Remote Execution server operating parameters during execution with the MODIFY command. All but MAXCONN, IPV6, and SECLABEL can be changed.

Step 5: Create a user exit routine (optional)

The TSO Remote Execution server enables you to provide an optional user exit routine. You can use this routine to alter the JOB and EXEC statement parameters. In addition, you can use this routine to provide JES control statements to meet installation-specific requirements, such as updating accounting information before the submission of the TSO batch job.

On entry to the exit, register 1 points to the following parameter list:

Offset
Description

+0
A pointer to an AF_INET or AF_INET6 address structure in the following format:

Offset
Description

+0
Address family AF_INET or AF_INET6 (2 bytes)

+2
Server port (2 bytes)

+4
Client AF_INET or AF_INET6 IP address (4 or 16 bytes)

+4
A pointer to a 1024-byte string terminated by the null character (0x00), which contains the JOB statement parameters. The exit routine can modify this string to meet installation-specific requirements. On entry, the string is set to the following string:

```
userid,USER=userid,[PASSWORD=passwd,]MSGCLASS=msgclass[,SECLABEL=seclabel]
```

The *userid* and *passwd* values are set to the user ID and password that the client provides, and the *msgclass* value is the message class of the Remote Execution server cataloged procedure. The PASSWORD parameter is present only when the client provides a password. The server passes the SECLABEL parameter only when the user ID is defined with a security label. The following example shows a modified string:

```
userid,USER=USER1,PASSWORD=USERXX1,MSGCLASS=E,LINES=999999,NOTIFY=SYSADM
```

+8
A pointer to a 256-byte string terminated by the null character (0x00), which contains the EXEC statement parameters. The exit routine can modify this string to meet installation-specific requirements. On entry, the string contains the EXEC statement for the procedure that is specified on the TSOPROC parameter of the Remote Execution server, or the default IKJACCNT procedure if the TSOPROC parameter is not specified. The following example shows a modified string:

```
IKJACCNT,REGION=4K,TIME=300
```

+12
A pointer to a 256-byte string terminated by the null character (0x00), which the exit can use to provide JES control statements. On entry, the string is set to 0x00. The following example shows a modified string:

```
/*JOBPARM  SYSAFF=SYS5
```

Exit return codes:

- 0

  A return code of 0 indicates to the server to continue processing the request.

- Nonzero

  A nonzero return code indicates to the server to send the job control buffer (offset +4) back to the client and close the connection.

Rules:

- Code the exit with the AMODE(31) and RMODE(ANY) attributes to ensure proper addressability of the input parameters.
- The exit must terminate all modified strings with the null character (0x00).

- If the server is IPv6 enabled and an IPv4 client connects, the IP address is the 4-byte IPv4 address and not the IPv4-mapped IPv6 address.
- If the exit provides JES control statements, the server inserts the control statements into the JCL stream following the JOB statement and before the EXEC statement without any parsing. If more than one control statement is present, the exit must ensure proper JCL line separation by including the NL character or CRLF character as necessary.
- The server inserts the EXEC parameters into the JCL stream without parsing. If the exit modifies the EXEC parameters, it must ensure proper JCL line separation by including the NL character or CRLF character as necessary.

Guidelines:

- For RSH commands without passwords, the PASSWORD parameter is not present.
- The `userid` in the first positional parameter of the JOB statement parameters can be processed by installation-written JES exits.
- The modified parameters are submitted as a TSO batch job.

Tip: The exit can have the server reject a connection by setting a nonzero return code. The exit can replace the JOB statement buffer contents with message text to be returned to the client. The first byte of the buffer can be set to 0x00 to reject the connection without sending a message to the client.

The user exit is shipped as a sample in the RXUEXIT member of the SEZAINST data set. For more information about this sample, see the REXEC topic in z/OS Communications Server: IP Configuration Reference.

Step 6: Permit access to JESSPOOL files

If the SAF resource class JESSPOOL is defined, ALTER access is required to access output files. Alternatively, use a jobname of *prefix**, where *prefix* is defined in RXSERVE.

Configuring the z/OS UNIX Remote Execution servers

This topic describes the z/OS UNIX files used by z/OS UNIX REXECD and RSHD.

Files for z/OS UNIX REXECD

Note: The user ID associated with the daemon in /etc/inetd.conf requires superuser authority. See z/OS UNIX System Services Planning for a description of the kinds of authority defined for daemons.

The files used by z/OS UNIX REXECD and their locations in the z/OS UNIX file system are as follows:

/etc/services
The ports for each application are defined here.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file.

/etc/inetd.conf
The configuration parameters for all applications started by inetd are defined in this file.

/usr/sbin/orexecd
The server.

If BPX.DAEMON is specified, then the sticky bit must be set on, and /usr/sbin/orexecd, and orexecd can be in an authorized MVS data set.

/usr/lib/nls/msg/C/rexdmsg.cat
The message catalog used by the z/OS UNIX REXECD server.

Note: This is not an actual member at this location, but it is a symbolic link to the part in /usr/lpp/tcpip/nls/msg/C/*.

Where the server looks for the message catalog (rexdmsg.cat) depends on the value of NLSPATH and LANG environment variables. If you want to store the msg.cats elsewhere, you need to change the NLSPATH or the LANG environment variables. If rexdmsg.cat does not exist, by default, the software uses the messages hard-coded within the software. These messages duplicate the English message catalog that is shipped with the product.

Files for z/OS UNIX RSHD

The files used by z/OS UNIX RSHD and their locations in the z/OS UNIX file system are as follows:

/etc/services
The ports for each application are defined here.

/etc/syslog.conf
The configuration parameters for usage of syslogd are defined in this file.

/etc/inetd.conf
The configuration parameters for all applications started by inetd are defined in this file.

/usr/sbin/orshd
The server.

If BPX.DAEMON is specified, the sticky bit must be set on, and /usr/sbin/orshd, and orshd can be in an authorized MVS data set.

/usr/sbin/ruserok
An optional user exit that will authenticate users logging into the z/OS UNIX RSHD server with a null password. See "Setting up the z/OS UNIX RSHD installation exit" on page 1314 for more information.

Note: This exit is required to allow support for null passwords with RSH.

/usr/lib/nls/msg/C/rshdmsg.cat
The message catalog associated with the z/OS UNIX RSHD client is stored here. If this file does not exist, by default, the software uses the messages hard-coded within the software. These messages duplicate the English message catalog that is shipped with the product.

Note: The message catalog is not actually stored here. This is a symbolic link, and the actual member is in /usr/lpp/tcpip/nls/msg/C/*.

Considerations for Kerberos authentication

If you plan to use Kerberos authentication with z/OS UNIX RSHD, you need to ensure that the user ID under which inetd runs has permission to the ICSF callable services that Kerberos needs. Specifically, if the CSFSERV class is active, ensure that the user ID of inetd has read access to the following ICSF SAF resources:

CSF1TRC, CSF1SKE, CSF1SKD, CSF1TRD, and CSFOWH

On z/OS, Kerberos is implemented by Security Server. See z/OS Integrated Security Services Network Authentication Service Administration for more information.

Setting up the z/OS UNIX RSHD installation exit

When the `-r` option is enabled, if there is no password specified on the RSH command from the client, z/OS UNIX RSHD will drive the installation exit. When the installation exit is driven, RSHD looks for a program in /usr/sbin named ruserok. This is the only name that it will look for. If /usr/sbin/ruserok is not found, the request will fail.

When the z/OS UNIX RSHD server invokes /user/sbin/ruserok, it will pass parameters in the following order:

1. Host name or the host IP address
2. Local user's UID
3. Remote user ID

4. Local user ID

If z/OS UNIX RSHD receives a return code of zero from the installation exit, z/OS UNIX RSHD continues. Any nonzero return code from the installation exit will cause RSHD to issue message EZYRS25E to the client and terminate all connections. The following code fragment can be used as an example to begin building a working ruserok installation exit:

```
int main(argc, argv)
     int argc;
     char *argv[];
  char *rhost1;    /* "hostname" or "hostname.domain" of client
                      obtained by caller:
                      gethostbyaddr(getpeername()) or the host
                      ip address used by the gethostbyaddr if
                      it failed to return a "hostname" */
  int locuid;      /* uid of the user name on local system */
  char *cliuname;  /* user name on client's system */
  char *servuname; /* user name on this (server's) system */
  int rc = 4;

  rhost1 = argv[1];
  locuid = atoi(argv[2]);
  cliuname = argv[3];
  servuname = argv[4];
  .
  <authenticate user and set rc=0 if valid>
  .
  return(rc);
```

Configuring TSO and z/OS UNIX Remote Execution servers to use the same port

Because the remote execution servers are generic servers, they attempt to bind to INADDR_ANY when they are started. This allows them to listen on all defined IP addresses. However, this prevents both the TSO and z/OS UNIX Remote Execution servers from listening on the same port, and one of the servers would have to use a nonstandard port. Using the BIND parameter on the PORT reservation statement in the TCPIP profile data set allows both the TSO and z/OS UNIX Remote Execution servers to bind to the same ports using different IP addresses. The following steps illustrate how this can be done. For more information on the PORT reservation statement, see z/OS Communications Server: IP Configuration Reference.

1. Define a VIPA address to the TCPIP profile data set. This example shows a static VIPA address, but either a static or dynamic VIPA can be used.

   ```
   DEVICE VIPAD1 VIRTUAL 0
   LINK   VIPA1  VIRTUAL 0 VIPAD1

   HOME
   134.134.134.36       VIPA1
   ```

2. Add PORT statements to the TCP/IP profile for both the TSO and z/OS UNIX Remote Execution servers. One of the servers will bind to the VIPA address. The other can bind to INADDR_ANY by not specifying the BIND parameter. In this example, the z/OS UNIX Remote Execution servers will bind to the VIPA address. Also update the /etc/services file so that exec uses 512 and shell uses 514.

   ```
   512 TCP OMVS BIND 134.134.134.36 ; z/OS Unix REXECD
   514 TCP OMVS BIND 134.134.134.36 ; z/OS Unix RSHD
   512 TCP RXSERVE                  ; TSO REXECD
   514 TCP RXSERVE                  ; TSO RSHD
   ```

   It is important that the server with the BIND parameter is listed before the one without the BIND parameter. This setup directs all requests to ports 512 or 514 with a destination IP address of 134.134.134.36 to the z/OS UNIX Remote Execution servers. Requests to ports 512 or 514 with a local destination IP address that is not 134.134.134.36 are directed to the TSO Remote Execution server.

To verify this setup:

1. Start the stack with the TCP/IP profile changes described above and start RXSERVE and INETD.

   Note: INETD listens for the REXEC and RSH servers under z/OS UNIX.

2. Issue NETSTAT and it should show that both the REXEC servers are listening on port 512 and both RSH servers are listening on port 514. INETD, which listens for the z/OS UNIX Remote Execution servers, listens only on the VIPA address.

```
MVS TCP/IP NETSTAT CS V1R2       TCPIP NAME: TCPCS            21:34:41
User Id  Conn     Local Socket           Foreign Socket         State
-------  ----     ------------           --------------         -----
INETDCS1 0000000D 134.134.134.36..514    0.0.0.0..0             Listen
INETDCS1 0000000E 134.134.134.36..512    0.0.0.0..0             Listen
RXSERVE  00000019 0.0.0.0..514           0.0.0.0..0             Listen
RXSERVE  00000018 0.0.0.0..512           0.0.0.0..0             Listen
```

Chapter 30. Express logon services with the Digital Certificate Access Server

The Digital Certificate Access Server (DCAS) is a z/OS TCP/IP server that you can use for enhanced logon solutions for z/OS applications. DCAS is part of several IBM express logon solutions, such as Express Logon Feature (ELF) and Web Express Logon (WEL), that work with IBM's host emulation products. DCAS provides a service for these solutions (with ELF, DCAS is needed only for the 3-tier solution) by providing them with z/OS user IDs and PassTickets for logon to host applications. A PassTicket is like a password. In addition, DCAS also provides an open interface and can be used by third-party express logon solutions to obtain user IDs and PassTickets for logon to z/OS applications. Note that DCAS does not support IBM Multi-Factor Authentication for z/OS.

Express Logon Feature

Express Logon Feature (ELF) is an enhanced logon solution that is provided by IBM host access products Host on Demand (HoD) and Personal Communications. ELF is also referred to as certificate-based logon. ELF enables you to log on to host applications using an X.509 certificate for authentication. For more information, see “What DCAS provides” on page 1317 and Appendix C, “Express Logon Feature,” on page 1335.

Web Express Logon

Web Express Logon (WEL) is a single-signon solution that is provided by IBM host access products Host on Demand (HoD) and Host Access Transformation Services (HATS). WEL enables users that are already authenticated, by Web-based logon for example, to log on to their host-based applications without having to reauthenticate by entering a user ID and password. In this case, the host access product communicates with the DCAS server to obtain a PassTicket. For more information about HoD, see http://www.ibm.com/software/rational/products/hostondemand/. For more information about HATS, see http://www.ibm.com/software/awdtools/hats/.

Using the DCAS server interface for your logon solutions

DCAS provides an open interface for clients to connect using TCP/IP to obtain information that can be used in providing enhanced host logon solutions. For details on interfacing with DCAS, see z/OS Communications Server: IP Programmer's Guide and Reference.

What DCAS provides

In general, DCAS provides a service for returning a PassTicket. A PassTicket is like a password and can be used to log on to z/OS applications. RACF provides PassTicket support using the PTKTDATA class. To understand PassTickets and using the secured signon function, see z/OS Security Server RACF Security Administrator's Guide. The type of information DCAS returns depends upon the type of information requested by the client. Also, DCAS configuration controls what type of information is allowed to be provided.

DCAS provides two types of information:

- Certificate-based

  Given an x.509 certificate and an application ID, DCAS returns the user ID that has been mapped to the certificate in RACF and a PassTicket. This can be used by logon services that want to provide certificate-based logons. In this case, the certificate provided must be associated with a valid user ID in RACF. For information on using RACDCERT to administer certificates, see z/OS Security Server RACF

© Copyright IBM Corp. 2000, 2022

Security Administrator's Guide. This support is used by IBM's Express Logon Feature (ELF) for the 3-tier solution.

- User ID-based

  Given a user ID and an application ID, DCAS returns a PassTicket. In this case, the end-user should already have been authenticated using a method such as Web-based sign on, and the logon solution provider must ensure this authentication prior to requesting the PassTicket. This support is used by IBM's Web Express Logon (WEL).

Requirements:

- You must configure the DCAS server. For details on configuring DCAS, see z/OS Communications Server: IP Configuration Reference. The DCAS server must be configured to support the IBM logon solution that you want, or when using it for a general or third-party solution, must be configured to return the information that the client requires. This requires that the system administrator that is configuring DCAS and the logon service provider using DCAS work together. The SERVERTYPE configuration statement in the DCAS server profile determines the type of information that DCAS can provide to connecting clients and services (certificate-based, user ID-based, or both).
- DCAS requires that all clients must connect by using TLS/SSL and that the client must be authenticated. DCAS uses AT-TLS for TLS/SSL. For more information, see "Customizing DCAS for TLS/SSL" on page 1318. Review the CLIENTAUTH parameter in the DCAS configuration profile for determining the level of DCAS client authentication. To understand how to configure SSL key rings and certificates for DCAS, see Appendix B, "TLS/SSL security," on page 1329.
- For all supported configurations, DCAS provides a PassTicket for z/OS applications that are targets for express logon. For these applications, a PassTicket data profile must be defined. RACF provides PassTicket support using the PTKTDATA class.

Customizing DCAS for TLS/SSL

Customize DCAS to implement TLS security by using AT-TLS policies.

Before you begin

Ensure that you have the following setup or configuration:

- A valid DCAS configuration file must be available for DCAS to start. The following list shows the search order of DCAS configuration file:
  1. DCAS_CONFIG_FILE environment variable
  2. /etc/dcas.conf
  3. tsouserid.DCAS.CONF
  4. TCPIP.DCAS.CONF
- DCAS uses AT-TLS for its TLS/SSL support. For more information about setting up AT-TLS for DCAS, see Steps for customizing DCAS for TLS/SSL and Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121. For more information about configuring key rings and certificates for DCAS, see Appendix B, "TLS/SSL security," on page 1329.

Requirement: To use AT-TLS, you must configure the Policy Agent and enable the TCP/IP stack for AT-TLS. To configure AT-TLS, see Chapter 20, "Application Transparent Transport Layer Security data protection," on page 1121.

Procedure

Complete the following steps to customize DCAS to implement TLS security by using AT-TLS policies.

1. Decide the level of authentication that you use for TLS sessions:
   - Server authentication only
   - Client authentication level 1

- Client authentication level 2
- Client authentication level 3

For more information about server authentication and client authentication, see “Secure Socket Layer overview” on page 1329.

2. Create the server key ring database and add the certificates that you need to the server key ring database. For more information about server authentication, see “Server authentication ” on page 1329.
   - For server authentication, because every TLS session handshake includes server authentication, you must add a certificate for this server to the server key ring database. If a server certificate is self-signed, you must also export this certificate to the key ring databases of the clients that log in by using TLS. If a server certificate is signed by a certificate authority (CA), this CA certificate, instead of the server certificate, must be in the client key ring databases. For more information about server authentication, see “Server authentication ” on page 1329.
   - For client authentication, if you use client authentication and self-signed certificates, you must import the client certificates into the server key ring database. If a client certificate is signed by a certificate authority (CA), this CA certificate, instead of the client certificate, must be in the server key ring database. For more information, see “Client authentication” on page 1330.
3. You must configure the AT-TLS policies for DCAS. For more information about configuring AT-TLS, see “Configuring the server system” on page 1126.
   - You must complete the following configuration for AT-TLS policies:
     - Code the TTLSEnvironmentAdvancedParms statement with the ApplicationControlled parameter ON in the policy configuration file. The ApplicatonControlled parameter enables DCAS to start TLS security on a connection.
     - Set the HandshakeRole parameter to ServerWithClientAuth on the TTLSEnvironmentAction statement in the policy configuration file.
     - Code the same level of client authentication, the CLIENTAUTH parameter, in the DCAS configuration file and AT-TLS policies, the ClientAuthType parameter.
     - The TTLSRule statement with the LocalPortRange parameter and the LocalAddr parameter must include the DCAS configured values for PORT and IPADDR.
   - Configure TTLSCipherParms statement in the policy configuration file to use ciphers.
   - See Digital Certificate access server (DCAS) configuration file keywords and parameters in z/OS Communications Server: IP Configuration Reference for detailed information.

Table 73. Required AT-TLS policies when DCAS TLSMECHANISM is set to ATTLS

DCAS Configuration Value	**AT-TLS Policy value (policy configuration file)**
	Direction Inbound
	TTLSEnabled On
	ApplicationControlled ON
	HandshakeRole ServerWithClientAuth
CLIENTAUTH LOCAL2 (default)	ClientAuthType SAFCHECK
CLIENTAUTH LOCAL1	ClientAuthType Required (default)
PORT <portValue>	LocalPortRange <include portValue>
IPADDR <inaddrValue>	LocalAddr <inlcude inaddrValue>

The following Policy Agent AT-TLS configuration sample shows the required policy configuration statements for AT-TLS:

```
TTLSGroupAction secure_DCAS_group
{
      TTLSEnabled On
}

TTLSEnvironmentAction secure_DCAS_env
{
      TTLSKeyringParms
      {
            Keyring DCAS/server-keyring-database
      }

      HandshakeRole ServerWithClientAuth

      TTLSEnvironmentAdvancedParms
      {
            ApplicationControlled On
            ClientAuthType SAFCHECK            # Used with CLIENTAUTH Local2
            TLSv1   Off
            TLSv1.1 On
            TLSv1.2 On
            TLSv1.3 On
      }

      TTLSCipherParmsRef DCAS_ciphers   # Used to customize ciphersuites for DCAS
}

TTLSCipherParms DCAS_ciphers
{
      # Sample ciphers. Should be customized!
      V3CipherSuites TLS_RSA_WITH_AES_256_CBC_SHA
      V3CipherSuites TLS_DHE_DSS_WITH_AES_256_CBC_SHA
      V3CipherSuites TLS_DHE_RSA_WITH_AES_256_CBC_SHA
      V3CipherSuites TLS_RSA_WITH_AES_128_CBC_SHA
      V3CipherSuites TLS_DHE_DSS_WITH_AES_128_CBC_SHA
      V3CipherSuites TLS_DHE_RSA_WITH_AES_128_CBC_SHA
      V3CipherSuites TLS_AES_128_GCM_SHA256
      V3CipherSuites TLS_AES_256_GCM_SHA384
}

TTLSRule secure_DCAS_rule
{
      LocalPortRange 8990     # This should be set to the port DCAS is
                              # listening on (PORT keyword in DCAS configuration file)
      Direction Inbound
      TTLSGroupActionRef secure_DCAS_group
      TTLSEnvironmentActionRef secure_DCAS_env
}
```

Tip:

- You can configure many other TLS settings in the AT-TLS policy.
- Default AT-TLS policies for DCAS are defined in the IBM Configuration Assistant for z/OS Communications Server.

When you define your AT-TLS policy rules for DCAS, you will need to decide on a variety of important TLS settings, such as the acceptable TLS protocol versions and the acceptable cipher suites. If need be, work with your network security administrator to determine the proper values for these settings.

4. Configure your security product for TLS. For more information, see Appendix B, “TLS/SSL security,” on page 1329.

Transport Layer Security (TLS) terms

The following terms apply to Transport Layer Security (TLS).

Partner authentication, server authentication, client authentication
Indicates an algorithm that is applied to verify the authenticity of the TLS/SSL partner. This is done by using X.509 digital certificates. Server authentication indicates that the server authenticates itself to the client. Client authentication indicates that the client authenticates itself to the server after the server authentication completes.

Data authentication protection, data authentication
Indicates that an algorithm is applied to the data that is being transferred. This algorithm modifies the data so that the receiving program can verify whether the data is originated from the expected sender.

Data integrity protection
Indicates that an algorithm is applied to the data that is being transferred. This algorithm modifies the data so that the receiving program can verify whether the data was not modified or changed by others during the transfer.

Privacy protection, encryption
Indicates that an algorithm is applied to the data that is being transferred. This algorithm encrypts or scrambles the data so that only the receiving program can use a special key to decrypt or unscramble the data to its original format.

The original data cannot be seen or interpreted when the data is being transferred.

Raw
Indicates that the data is transferred without being modified by any encryption, data authentication, or data integrity algorithms.

Cipher suite
A collection of partner authentication, data authentication, data integrity, and encryption algorithms, which are used in combination for a given TLS/SSL session.

Note: This term does not indicate the algorithms that are used and does not indicate that the data is encrypted.

Chapter 31. Miscellaneous server

The Miscellaneous (MISC) server is a server that can be used to test and debug applications.

The MISC server supports the 3 protocols described in RFCs 862, 863, and 864:

- Discard
- Echo
- Character Generator

Discard protocol

The MISC server throws away any data it receives. A TCP-based server listens for TCP connections on TCP port 9. If a connection is established, the data is discarded and no response is sent. A UDP-based server listens for UDP datagrams on UDP port 9. When a datagram is received, it is discarded and no response is sent.

Echo protocol

The MISC server returns to the originating application any data that it receives. A TCP-based server listens for TCP connections on TCP port 7. After a connection is established, any data that is received is sent back to the originating application. A UDP-based server listens for UDP datagrams on UDP port 7. When a datagram is received, the data it contained is sent back as an answering datagram.

Character generator protocol

The MISC server sends a repetitive stream of character data without regard to its content. A TCP-based server listens for TCP connections on TCP port 19. When a connection is established, a stream of data is sent to the connecting application. Any data that is received is thrown away. A UDP-based server listens for UDP datagrams on port 19. When a datagram is received, an answering datagram is sent that contains a random number (between 0 and 512) of characters. The data in the received datagram is ignored.

The data that is generated follows an ordered sequence. It repeats a pattern of 94 printable ASCII characters in a ring, so that character number 0 follows character number 94.

The following example shows the repeated pattern.

```
!"#$%&'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefgh
"#$%&'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghi
#$%&'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghij
$%&'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijk
%&'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijkl
&'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklm
'()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmn
()*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmno
)*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnop
*└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopq
└,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqr
,-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrs
-./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrst
./0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstu
/0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuv
0123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvw
123456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwx
23456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxy
3456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz
456789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{
56789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|
6789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}
789:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~
89:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~
9:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~ !
:;←=→?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]¬_`abcdefghijklmnopqrstuvwxyz{|}~ !"
```

© Copyright IBM Corp. 2000, 2022

Configuring the MISC server

Procedure

1. Specify AUTOLOG and PORT statements in *hlq*.PROFILE.TCPIP.
2. Update the MISC server cataloged procedure (MISCSERV).

Step 1: Configuring PROFILE.TCPIP for the MISC server

To allow the MISC server to start automatically when TCPIP is initialized, include the member name of the MISC server cataloged procedure in the AUTOLOG statement in the *hlq*.PROFILE.TCPIP.

```
    AUTOLOG
      MISCSERV
    ENDAUTOLOG
```

The AUTOLOG entry in *hlq*.PROFILE.TCPIP is optional. You can choose to start the MISC server manually, when it is needed, using the START command:

```
    START MISCSERV
```

The MISC server requires ports 7, 9, and 19 for both TCP and UDP. To ensure that these ports are reserved for the MISC server, verify that they are assigned to the member containing the MISC server cataloged procedure in the PORT statement in PROFILE.TCPIP.

```
PORT
    7 UDP MISCSERV
    7 TCP MISCSERV
    9 UDP MISCSERV
    9 TCP MISCSERV
   19 UDP MISCSERV
   19 TCP MISCSERV
```

For more information on these statements, see the z/OS Communications Server: IP Configuration Reference.

Step 2: Updating the MISC server cataloged procedure

Update the MISC server cataloged procedure by copying the sample in SEZAINST(MISCSERV) to your system or recognized PROCLIB and modifying the parameters and data set names to suit your local conditions.

MISC server cataloged procedure (MISCSERV)

```
//MISCSERV PROC MODULE=MISCSRV,PARMS=''
//*
//* TCP/IP for MVS
//* SMP/E Distribution Name: SEZAINST(MISCSERV)
//*
//*        Licensed Materials - Program Property of IBM.
//*        "Restricted Materials of IBM"
//*        5694-A01 (C) COPYRIGHT IBM CORP. 1994, 2003
//*        Status = CSV1R5
//*        Distribution library SEZAINST(MISCSERV)
//*
//MISCSERV EXEC PGM=&MODULE,
//           REGION=4096K,TIME=1440,
//           PARM='&PARMS'
//*
//*        The C runtime libraries should be in the system's link list
//*        or add them to the STEPLIB definition here. If you add
//*        them to STEPLIB, they must be APF authorized.  Change
//*        the name as appropriate for your installation.
//*
//STEPLIB  DD DISP=SHR,
//         DSN=TCPIP.SEZATCP
```

```
//SYSPRINT DD SYSOUT=*
//SYSIN    DD DUMMY
//SYSMDUMP DD SYSOUT=*
//*
//*        MSMISCSR identifies an optional data set for NLS support.
//*        It specifies the MISC server message repository.
//*
//*MSMISCSR DD DISP=SHR,
//*         DSN=TCPIP.SEZAINST(MSMISCSR)
//*
//*       SYSTCPD explicitly identifies which data set is to be
//*       used to obtain the parameters defined by TCPIP.DATA
//*       when no GLOBALTCPIPDATA statement is configured.
//*       See the IP Configuration Guide for information on
//*       the TCPIP.DATA search order.
//*       The data set can be any sequential data set or a member of
//*       a partitioned data set (PDS).
//*
//SYSTCPD  DD DISP=SHR,
//         DSN=TCPIP.SEZAINST(TCPDATA)
```

Specifying the MISC server parameters

The MISC server generates periodic messages whenever a client sends data to ports 7, 9, or 19. If this server runs continually for a long period of time, considerable amounts of spool space can be consumed. Therefore, the MISC server has all tracing turned off by default.

You can enable the trace options for any of the three MISC server protocols using the PARMS= parameter on the PROC statement of the cataloged procedure. These options will be in effect when the server starts.

TRACE
Turns on tracing for any of the specified protocols and must be followed by one or more of these three keywords:

ECho
Specifies tracing for the echo protocol on port 7.

DIscard
Specifies tracing for the discard protocol on port 9.

CHargen
Specifies tracing for the character generator protocol on port 19.

DEbug
Specifies tracing for problem determination.

For example, the following statement turns tracing on for the echo and discard protocols.

```
//MISCSERV PROC MODULE=MISCSRV,PARMS='TRACE ECHO DISCARD'
```

Appendix A. Setting up the InetD configuration file

The Internet Daemon (InetD) is a generic listener program that is used by such servers as the z/OS UNIX telnet server and the z/OS UNIX rexec server. Other servers such as the z/OS UNIX ftp server have their own listener program and do not use InetD. For more information about generic servers and sample InetD started procedure JCL, see "Generic servers in a CINET environment" on page 48.

The inetd.conf file is an example of the user configuration file. This file is stored in the `/etc` directory and is referenced as a start parameter in the InetD started procedure JCL. The inet.conf file can also be stored as an MVS data set. Ensure that you use configuration statements like those statements in the following example to enable the InetD services that are required on your system:

```
#====================================================================
# service | socket | protocol | wait/ | user | server  | server program
# name    | type   |          | nowait|      | program |   arguments
#====================================================================
#
shell    stream    tcp        nowait  OMVSKERN /usr/sbin/orshd orshd -l
exec     stream    tcp        nowait  OMVSKERN /usr/sbin/orexecd orexecd -lv
otelnet  stream    tcp        nowait  OMVSKERN /usr/sbin/otelnetd otelnetd
# Add the following line to enable Kerberos for orshd
kshell   stream    tcp        nowait  OMVSKERN /usr/sbin/orshd orshd -l -k KRB5
```

Figure 158. Adding applications to /etc/inetd.conf

If the rshd, rexecd, or otelnetd service is to support IPv6 clients, then *tcp6* should be specified instead of *tcp*. Kerberos is not supported for IPv6-enabled services, such as z/OS UNIX Telnet, z/OS UNIX rsh, and z/OS UNIX rexec.

For IPv4 connection partners, the terminal ID passed from InetD to RACF (or an equivalent security program) is an 8-byte hexadecimal character string containing an IPv4 address. For example, the IP address 163.97.227.17 is translated to X'A361E311'. RACF interprets this as a terminal logon address and rejects it if it is not previously defined.

For IPv6 connection partners, only IPv4-mapped IPv6 addresses are handled in this way. The IPv4 address portion of the IPv6 address is placed in the terminal ID for RACF validation. No other IPv6 address format is supported through terminal ID RACF validation.

To establish a relationship between the servers defined in the /etc/inetd.conf file and specific port numbers in the z/OS UNIX environment, ensure that statements have been added to ETC.SERVICES for each of these servers. See the sample ETC.SERVICES installed in the /usr/lpp/tcpip/samples/services directory for how to specify ETC.SERVICES statements for these servers.

The traces for both the z/OS UNIX rexec and rsh servers are enabled through options in the InetD configuration file (/etc/inetd.conf):

```
#====================================================================
# service | socket | protocol | wait/ | user | server  | server program
# name    | type   |          | nowait|      | program |   arguments
#====================================================================
#
shell      stream    tcp       nowait OMVSKERN /usr/sbin/orshd orshd -d  1
exec       stream    tcp       nowait OMVSKERN /usr/sbin/orexecd orexecd -d  2
```

Figure 159. Setting traces in /etc/inetd.conf

The traces are turned on for both servers by passing a `-d` argument to the server programs. **1** is the RSHD server and **2** is the REXECD server. All commands executed after the debug flags have been turned on in the InetD configuration file and the InetD server has reread the file will produce trace output.

© Copyright IBM Corp. 2000, 2022

The trace is written in formatted form to the syslogd facility name daemon with a priority of debug. The trace data can be routed to a file in your Hierarchical File System by specifying the following definition in your syslogd configuration file (/etc/syslogd.conf):

```
#
# All ftp, rexecd, rshd
# debug messages (and above
# priority messages) go
# to server.debug.a
#
daemon.debug              /tmp/syslogd/server.debug.a
```

In this example, the trace data is written to /tmp/syslogd/server.debug.a in your Hierarchical File System. For more information on syslogd, see "Logging of system messages" on page 30.

For more information about InetD, see z/OS UNIX System Services Planning or z/OS UNIX System Services Command Reference.

Appendix B. TLS/SSL security

The Transport Layer Security (TLS) protocol is defined by the Internet Engineering Task Force (IETF) in RFCs 2246, 4346, 5246, and 8446. TLS is based on the Secure Socket Layer (SSL) protocol.

This appendix is a reference for z/OS Communications Server applications that use TLS/SSL.

More information about the concepts of cryptography and SSL can be found at the following websites:

- http://httpd.apache.org/docs/current/ssl/ssl_intro.html
- http://www.verisign.com/ssl/ssl-information-center/how-ssl-security-works/index.html

For examples of digital certificate implementation scenarios, see z/OS Security Server RACF Security Administrator's Guide.

Secure Socket Layer overview

SSL provides data privacy and integrity as well as server and client authentication based upon a Public Key Infrastructure (PKI) method. PKI requires that the server organization generate a public key/private key pair that can be used during negotiations. PKI requires that data encrypted with the public key be decrypted by only the private key and that data encrypted with the private key be decrypted by only the public key. This is considered an asymmetric encryption method because different keys are used at each end of the secure connection. The Server sends its public key to the client when the client requests a connection.

The client and server encrypt SSL parameter negotiations using the PKI method of encryption. One of the most important items negotiated is the encryption algorithm to be used during data transmission. The algorithm chosen will be one that uses the same key at each end of the secure connection. This is known as a symmetric encryption method and is about 1000 times faster than the asymmetric PKI method used during SSL parameter negotiation. The encryption key used by the symmetric encryption method is created and exchanged during SSL negotiation protected by the PKI encryption method.

Some client-server connections support negotiations to determine if the client wants or supports SSL prior to beginning the SSL handshake. Most servers and clients can be configured to immediately start the SSL handshake process or to negotiate whether or not to perform the SSL handshake. See the security information for the appropriate server or client for information on whether negotiated TLS/SSL is supported and how it is implemented.

The SSL protocol begins with the handshake. During the handshake:

- Server authentication is done by the client.
- Optional client authentication is done by the server.
- An encryption algorithm and single encryption key are chosen to encrypt and decrypt session data between the client and server.

Server authentication

When using SSL to secure communications, the SSL authentication mechanism known as Server Authentication is used. This is the minimum amount of security provided by SSL and allows the client to validate that the Server is what it says it is.

To ensure that someone has not stolen the server's private and public keys and is pretending to be the server, the server sends additional information with the public key so the client can confirm the identity of the server. The complete package of information sent to the client is called a digital certificate which conforms to the X.509 standard.

This X.509 digital certificate includes, among other things, the Distinguished Name (DN) of the Server organization, the public key created by the server organization, the Distinguished Name of the

© Copyright IBM Corp. 2000, 2022

organization issuing the certificate, and the issuer's signature. The organization issuing the certificate may be a well-known Certificate Authority (CA) or you may issue (create) your own certificate, called a self-signed certificate.

To create a signature, the certificate issuer first generates a message digest from the owner's DN, the owner's public key, and the issuer's DN. The message digest is the result of hashing this information down to a small size, such as 128, 160, 256, or some other number of bits. The message digest result is unique for that information and is based on the hashing algorithm that is used. The message digest is encrypted with the issuer's private key to create the issuer's signature.

When the client receives the server certificate, the client must have the public key of the certificate signer. The public key is used to decrypt the message digest. The server certificate also contains the hashing algorithm used to create the message digest. The client uses the same algorithm to create another message digest using the Distinguished Names and public key information in the received server certificate. If this new message digest exactly matches the decrypted message digest (issuer's signature) created by the certificate issuer, the client can be assured that the certificate has not been altered. This method of authentication is dependent on the security of the private key that is used by the certificate issuer.

To conduct commercial business on the Internet, you should obtain a server certificate signed by a well-known Certificate Authority. Server certificates issued by a well-known CA gives the client high assurance that the server is authentic. Most client key rings have been primed with several well-known CA's certificates. That enables the client to authenticate a Server certificate signed by a well-known CA without having to first obtain the issuer's certificate which includes the public key. For relatively small, private networks within your own enterprise you can create your own self-signed server certificate. The only difference between a CA issued certificate and a self-signed certificate is the issuer's Distinguished Name and who's private key was used to encrypt the message digest. The client needs to use the correct public key to decrypt the message digest. The CA certificate containing the CA's public key is probably already in the client's key ring and it can be used to decrypt the CA signature (message digest). The self-signed certificate containing the organization's public key needs to be added to the client's list of signer certificates so the client can decrypt the signature (message digest) created when the self-signed certificate was created. Some client products allow the client to add the server certificate to its list of signer certificates when the server certificate is received during SSL negotiation. If the client is confident the certificate really came from the correct server, this is an easy way to add the certificate rather than getting a copy and adding it manually.

For server authentication to work, the server must have a private key and associated server certificate in the server key database file. The gskkyman utility or RACF Common Keyring support can be used to manage the keys and certificates needed for SSL support. If the gskkyman utility was used to create the key ring, a password stash file is also required.

SSL requires a server certificate as part of its server authentication process. The server certificate and the Certificate Authority certificates are stored in a key ring (also referred to as a key database). The server's key ring can be created using the gskkyman utility provided by the System Secure Socket Layer (System SSL) element of z/OS or by using RACF's certificate management support. The key ring is associated with a server or client using server or client specific statements.

Note: Global step-up type certificates are not supported by Telnet profile defined security if the client application sends the handshake complete message to the server before completing the second handshake. AT-TLS enabled Telnet does support global step-up type certificates.

Client authentication

Client authentication provides additional authentication and access control by checking client certificates at the server. This support prevents a client from obtaining a connection without an installation approved certificate.

The server authenticates the client by receiving the client's certificate during the SSL handshake and verifying the certificate is valid. Validation is done by the server the same way the client validates the server's certificate. The client sends a signed certificate to the server. System SSL at the server decrypts the signature (message digest) using the public key of the client certificate issuer found in the server key

database file. The server then creates a new message digest using the certificate's Distinguished Names and public key and compares the new message digest with the decrypted one. If they match, the server can be assured the client is authentic. Depending on where the client certificate is stored, up to three different levels of client authentication are available. See the security information for the appropriate server or client for setup details.

Level 1 authentication is performed by System SSL. The client passes an X.509 certificate to the Server as part of the SSL Handshake. To pass authentication, the Certificate Authority (CA) that signed the client certificate must be considered trusted by the server. That is, the certificate for the CA must be in the key ring used by the Server and designated as trusted. Note that the value of this option alone is based on which CAs are considered trusted. If the CA is a public CA and the certificate is in an easily obtained class, anyone can obtain such a certificate. In this case passing level 1 SSL Client Authentication does not provide much additional security unless coupled with the level 2 RACF support. If the CA is controlled by the enterprise, some level of access control is provided because the client that possesses such a certificate is at least known to the organization.

In addition to the checking done with the first level of client authentication support, *level 2* authentication requires that the client certificate is registered with RACF or another SAF-compliant security product and is mapped to a valid z/OS user ID. The client certificate that is received during the SSL handshake is used to query the security product to verify that the certificate maps to a user ID known to the system before connection negotiation. This level of support provides extra access control at the server and ensures that the user is known to have a valid user ID on the server host. Each server uses the returned user ID in a different way. To see how the user ID is used for a particular server, see the security information for the appropriate server or client. Level 1 authentication occurs before level 2 authentication.

Level 3 authentication provides, in addition to level 1 and level 2 support, the capability to restrict access to the server based on the user ID returned from RACF. In some cases a certificate may be valid and mapped to a user ID but should be valid for only one of several servers. The third level of control uses the SERVAUTH RACF class to restrict access to the server based on client user ID. If the SERVAUTH class is not active or the SERVAUTH profile for the server is not defined, it is assumed level 3 authentication is not requested. If the SERVAUTH class is active and the server profile is defined, a connection is accepted only if the requester's user ID associated with the client certificate is in the profile. Otherwise, the connection is dropped.

To enable Client Authentication for each server, use the following server-specific statements:

AUTHENTICATION LEVEL	**FTP**	**Telnet**	**DCAS**
	SECURE_LOGIN statement	**AT-TLS ClientAuthType setting**	**CLIENTAUTH statement**
AUTH LEVEL 1	REQUIRED	Required	LOCAL 1
AUTH LEVEL 2	VERIFY_USER	SAFCheck	LOCAL 2
AUTH LEVEL 3	VERIFY_USER	SAFCheck	LOCAL 2

Note: See Achieving the basic level of security for more information on the ClientAuthType parameter settings.

Level 1 client authentication is done by SSL using a gskkyman key ring or a RACF key ring. If the client certificate was issued by a well-known Certificate Authority, it is likely the CA certificate is already primed in the gskkyman key ring. The CA certificate is probably also in RACF. However, all CA certificates in RACF initially have a status of NOTRUST. The CA certificate must be set to TRUST and connected to the appropriate RACF key ring. If the certificate issuer (a CA or self-signed) is not part of the list of well-known CAs, the key ring must be primed with the signer certificate of the CA or the self-signed client certificate.

After Level 1 authentication is performed by SSL using either key ring, the certificate is passed to the server which accesses the RACF database for Level 2 and Level 3 authentication.

Encryption algorithms

After authentication occurs, the client and server must agree on a symmetric encryption method and generate a single encryption key to use for data encryption. The chosen key is exchanged by using the PKI method of encryption. After the symmetric encryption algorithm (such as AES) and a single encryption key are chosen, all data exchanges use this algorithm and key instead of the PKI method of encryption.

In an SSL-encrypted session, all data is encrypted with the symmetric encryption algorithm immediately before it is sent to the client. Data from the client is decrypted immediately after it is received. The encryption algorithm that is used for the connection depends on a combination of the following factors:

- The encryption algorithm list that the SSL subsystem supports
- The list that the server wants to use
- The encryption algorithms that the client requests

During the SSL handshake, the client sends a list of encryption algorithms it is able to use. The server submits its list and the SSL subsystem picks an algorithm that all parties support, giving preference to the order that the server specifies. If the server does not support any of the encryption algorithms that the client requests, the connection is closed. Several z/OS Communications Server applications use the SSL support that is provided by the System Secure Sockets Layer (System SSL) element of z/OS. The encryption algorithms that the servers and client support depend on the level of System SSL that is installed. The following encryption algorithms are supported by the base level of System SSL:

- NULL
- RC2 export
- RC4 export
- DES

The System SSL Level 3 feature is required for Triple DES, AES, ChaCha-Poly1305, and RC4 non-export (128 bit) encryption algorithms. The encryption algorithm list can be customized for the servers and client to a subset of the System SSL list. For specific server and client statements that are used to create the encryption list, see the security information for the appropriate server or client.

Encryption is provided by either hardware, if available, or software algorithms within System SSL or Integrated Cryptographic Services Facility (ICSF). There is no TCP profile definition to control whether the cryptographic hardware is used for secure connections. For information about cryptographic hardware usage within System SSL, see z/OS Cryptographic Services System SSL Programming. To use FIPS 140 mode, ICSF must be started and available to System SSL.

If hardware encryption is to be used, be sure that the RACF user ID associated with the server has read access to the RACF CSFSERV class resources. If ICSF is available but the server has not been given access to these resources, the SSL initialization may fail. The reason code is likely to be 4 (bad password) because System SSL will attempt to use the hardware encryption during processing of the key ring.

Enable CSFSERV resources

If hardware encryption and Integrated Cryptographic Services Facility (ICSF) are installed, system SSL verifies that the user ID that is associated with the server is permitted to use CSFSERV resources. The RACF administrator can permit the RACF user ID to use the CSFSERV resources:

```
PERMIT service-name CLASS(CSFSERV) ID(serverid) ACCESS(READ)
```

For information about the CSFSERV resources (service-names) that are accessed by System SSL, see z/OS Cryptographic Services System SSL Programming.

z/OS FTP users can either permit every FTP client user ID to these general resource profiles, or they can mark these profiles as delegated and permit only the FTP daemon user ID to the profiles.

In the following example, resource CSFENC in class CSFSERV is delegated, and only the FTP daemon user ID (FTPD for this example) needs to be permitted. Make these changes before you start FTPD.

Permit the FTP daemon to the resource:

```
PERMIT CSFENC CLASS(CSFSERV) ID(FTPD) ACCESS(READ)
```

Mark the resource profile as delegated:

```
RALTER CSFSERV CSFENC APPLDATA('RACF-DELEGATED')
```

Refresh the CSFSERV class:

```
SETROPTS RACLIST(CSFSERV) REFRESH
```

For more examples, see the EZARACF sample in SEZAINST. For more information about authorizing daemons to use delegated resources, see z/OS Security Server RACF Security Administrator's Guide.

The MAXLEN installation option for hardware cryptography determines the maximum length that can be used to encrypt and decrypt data by using ICSF. Set this option to 65527 or greater, which is the maximum TCP/IP packet size.

The System SSL GSKSRVR server provides the capability to determine whether cryptographic hardware is being used through its DISPLAY CRYPTO operator command (for example, f gsksrvr,d crypto). The System SSL GSKSRVR server is not automatically started. For more information about the SSL started task and setting up and using the GSKSRVR server, see z/OS Cryptographic Services System SSL Programming.

For more information about controlling who can use cryptographic keys and services, see z/OS Cryptographic Services ICSF Administrator's Guide.

Appendix C. Express Logon Feature

Users accessing SNA applications using Telnet clients such as Host On Demand are generally required to know the user ID and password for the application they want to access. The ID-and-password authentication process creates several potential problems. For example, users may forget their IDs and passwords. If they do forget, the passwords must be reset by a system administrator, a time-consuming process. On the other hand, writing down the IDs and passwords or sharing them with someone else creates a security risk, especially because passwords are usually valid for relatively long periods of time.

IBM's solution to these problems is the Express Logon Feature (ELF), a process which allows a user on a workstation with a Telnet client and an X.509 certificate to log on to an SNA application without entering an ID or password. The Express Logon Feature is supported on two-tier and three-tier network designs. The two-tier design uses the z/OS TN3270E Telnet server. The three-tier design uses a middle-tier Telnet server and a Digital Certificate Access Server (DCAS).

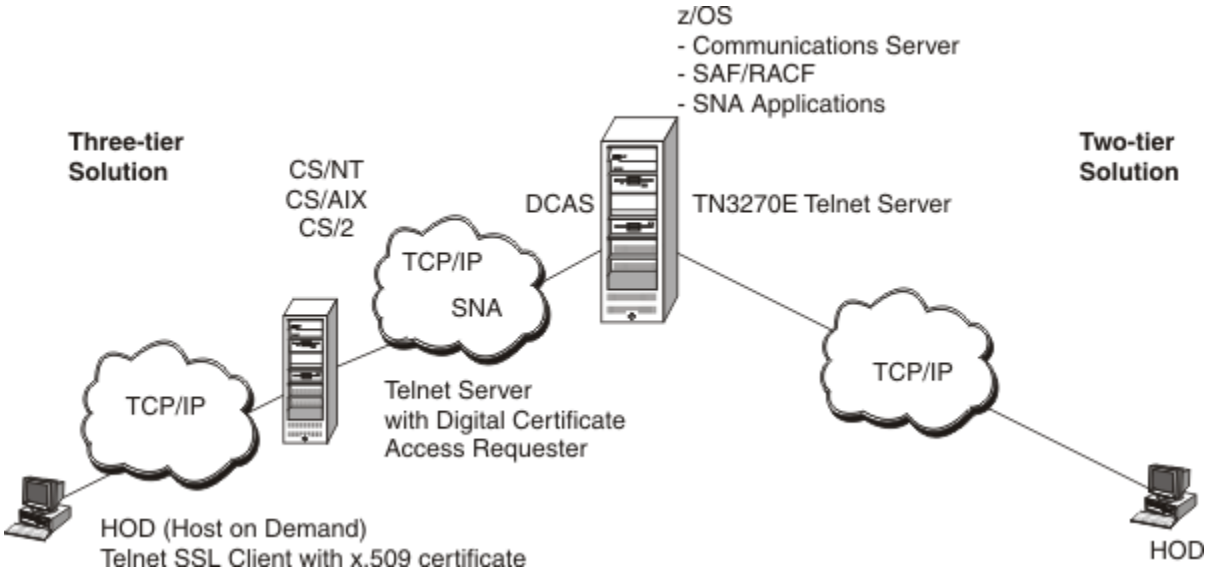


Figure 160. Express Logon network

Both network designs require a Telnet client emulator that supports:

- Transport Layer Security (TLS) or Secure Sockets Layer (SSL) connections with client authentication and an X.509 certificate. Using RACF services in z/OS, the client certificate must be associated with a valid user ID.
- The Express Logon Feature (ELF).

The two-tier design requires the z/OS TN3270E Telnet server with TLS/SSL, client authentication, and Express Logon functions turned on. See “Express Logon Feature” on page 643 for server setup information.

The three-tier design requires a middle-tier Telnet server that includes a Digital Certificate Access Requester (DCAR) and the z/OS Digital Certificate Access Server (DCAS). A middle-tier Telnet server, so called because it is not on the host, but rather between the Telnet client emulator and the host.

Note: The term *DCAR* is used to describe the part of the Telnet middle-tier server that supports the Express Logon Feature and communicates as a client with the DCAS. It is not separate from the Telnet middle-tier server. The term DCAR might not be used in other information that describes ELF but has been used here to simplify the description of this function.

A Digital Certificate Access Server (DCAS) exists on the host. DCAS uses RACF services to obtain a user ID that has been mapped to a digital certificate.

© Copyright IBM Corp. 2000, 2022

The host also provides RACF Secured Signon services, which the DCAS or the z/OS host TN3270E Telnet server uses to generate a PassTicket. A PassTicket is a RACF token similar to a password except that it is valid only for ten minutes. Additionally, the z/OS TN3270E Telnet server might use a Multi-Factor Authentication (MFA) token that was generated by IBM MFA for z/OS.

In a typical scenario, an ELF-enabled client wants to log on to a TSO application on the host.

- In the two-tier design, the user starts a secure connection with level 2 client authentication, which passes the client certificate to the TN3270E Telnet server. The TN3270E Telnet server uses either RACF MFA services to retrieve a user ID and an MFA token or uses RACF Secured Signon services to obtain a user ID and PassTicket.
- In the three-tier design, the user starts the Telnet connection to the middle-tier Telnet server. The client of the DCAS is the middle-tier Telnet server or DCAR, which attempts to log on to an SNA application for the client emulator. The DCAS receives a digital certificate from the DCAR and returns a user ID and PassTicket. Secure communication is used between the DCAS and the DCAR. The server recognizes that the client wants the Express Logon function and invokes the DCAR, which opens a secure connection with client authentication and passes the client's certificate and application name to the DCAS on the host. The DCAS uses RACF Secured Signon services to obtain a user ID and PassTicket, which the DCAS returns to the DCAR. The DCAR passes this information back to the middle-tier Telnet server.

In both cases the ELF-enabled client and server now have enough information to complete the logon to TSO. This occurs without the user ever having to enter a user ID or password.

Note: You can use RACF or any other SAF-compliant security product that supports MFA (for the two-tier design) or PassTickets with Express Logon.

Configuring RACF services for Express Logon

At a minimum, you must register all client certificates with RACF or create a RACF certificate name filter for the client certificates by using the RACDCERT command. This associates the certificates with the IDs of users who are attempting to log on. In the two-tier solution, the certificate is passed from the client to the TN3270E Telnet server. In the three-tier solution, the certificate is passed from the client to the middle-tier Telnet server, then to the DCAR, and then to the DCAS.

Configuring RACF and IBM Multi-Factor Authentication for z/OS (two-tier solution only)

To use IBM Multi-Factor Authentication (MFA) for z/OS, you must set up and configure the IBM MFA for z/OS product and also add appropriate MFA parameters to the RACF user profile for the users that need to log in with MFA tokens. For more information about IBM MFA for z/OS setup and configuration, see the following documentation:

- *IBM MFA V1R1: TouchToken, PassTicket, and Application Bypass Support* (An IBM Redpaper publication: IBM Form #: REDP-5386)
- z/OS Security Server RACF Security Administrator's Guide
- IBM Multi-Factor Authentication for z/OS Installation and Customization

Configuring RACF for PassTickets

You must also create a RACF PTKTDATA profile for each application ID the user is attempting to access. The PTKTDATA profile allows the DCAS or z/OS TN3270E Telnet server to obtain a PassTicket and user ID for the application. In the three-tier solution, the DCAS must pass the PassTicket and user ID back to the DCAR. For Host On Demand, the application ID part of the profile name must be the same as that configured in the Host On Demand Express Logon Application ID popup window. In most cases, the application name with which the user logs on will match the application ID portion of the RACF PTKTDATA class profile. However, for TSO and some other applications, the names and IDs may not match:

- If VTAM generic resources are used for TSO, define the application name portion on the RACF profile using the TCASGNAM defined in the TSOKEYxx, SYS1.PARMLIB member.
- If VTAM generic resources are not used, define the application name on the RACF profile as TSO.
- When configuring for TSO application logon, use the format TSO<SID> in the PassTicket profile, where SID is the SMF system ID defined in the SMFPRMxx member of SYS1.PARMLIB. (For example, if the SID

is 3390, you would type TSO3390 in the profile.) For details, see z/OS Security Server RACF Security Administrator's Guide.

For applications that allow shared user IDs (multiple users request access to the application simultaneously with the same user ID), you must specify the APPLDATA('NO REPLAY PROTECTION') option on the RDEFINE command in the PTKTDATA profile. This bypasses the default RACF protection against replay of PassTickets.

An example of configuring the Express Logon components

The following example describes, in general terms, how to set up and configure Express Logon by using the following products and components:

- IBM Host On Demand Telnet client
- z/OS TN3270E Telnet server (for two-tier solution)
- IBM Communications Server for Windows (middle-tier Telnet server for three-tier solution)
- DCAS - see Chapter 30, “Express logon services with the Digital Certificate Access Server,” on page 1317

For details on configuring each product, see the appropriate documentation for that product.

Configuring the Host On Demand Telnet client

To setup and configure the Host On Demand client, follow these steps:

1. For each application to which the user will log on, create a macro to record the logon screen of the application.

   The user plays this macro when displaying the logon screen on the client.
2. Enter the application ID in the application ID popup window.

   The application ID must be the same name specified on the z/OS for the application ID portion of the PTKTDATA profile. The ID in the profile in most cases must be the same as the application name.
3. Use the Host On Demand key-management utility to:
   a. Create a key database (key ring).
   b. Create a certificate request or generate a self-signed certificate and associate the certificate with the key ring.
   c. Use FTP to transmit the middle-tier Telnet certificate to the client workstation and store the server certificate in the key database of the client.
   d. Use FTP to transmit the Host On Demand Telnet client certificate to the middle-tier server and store in the SSL key database of the server.
4. Use FTP to transmit the client certificate to an MVS data set on the z/OS host. Use the RACF Certificate Services RACDCERT command to associate the certificate with a valid user ID.

Configuring the z/OS TN3270E Telnet server (two-tier solution)

Express Logon Feature requires TLS/SSL with level 2 client authentication functionality at the server. After that level of security is working, specify the EXPRESSLOGONMFA (to use Multi-Factor Authentication tokens) or EXPRESSLOGON (to use PassTickets) parameter statement to enable ELF in the z/OS TN3270E Telnet server.

Configuring the middle-tier Telnet server (IBM Communications Server for Windows example)

The middle-tier server is a Telnet server such as IBM Communications Server for Windows, that communicates with the Host On Demand client using a TLS/SSL connection with client authentication.

The middle-tier server DCAR also communicates with the DCAS on the host. The DCAS and DCAR communicate over a TCP/IP connection using TLS/SSL with client authentication.

To configure the Telnet server, follow these steps:

1. Use the Communications Server for Windows SNA Node Configuration panels to enable the Express Logon Feature and configure the DCAS server address and port.
2. Use the local key management utility to store the client certificate and the DCAS certificate in the local key ring:
   a. Create a key database file.
   b. Create a certificate request or generate a self-signed certificate and associate the certificate with the key ring.
   c. Store the client certificate and the DCAS certificate in the key ring of the server.
3. Use FTP to transmit the DCAR certificate to the z/OS host and use gskkyman or RACF Certificate Services to store the DCAR certificate in the DCAS key ring.

Appendix D. Using HCD

This information includes examples of panels that are used to define IQD channels and devices for z/OS Communications Server using HCD.

1. Select processors

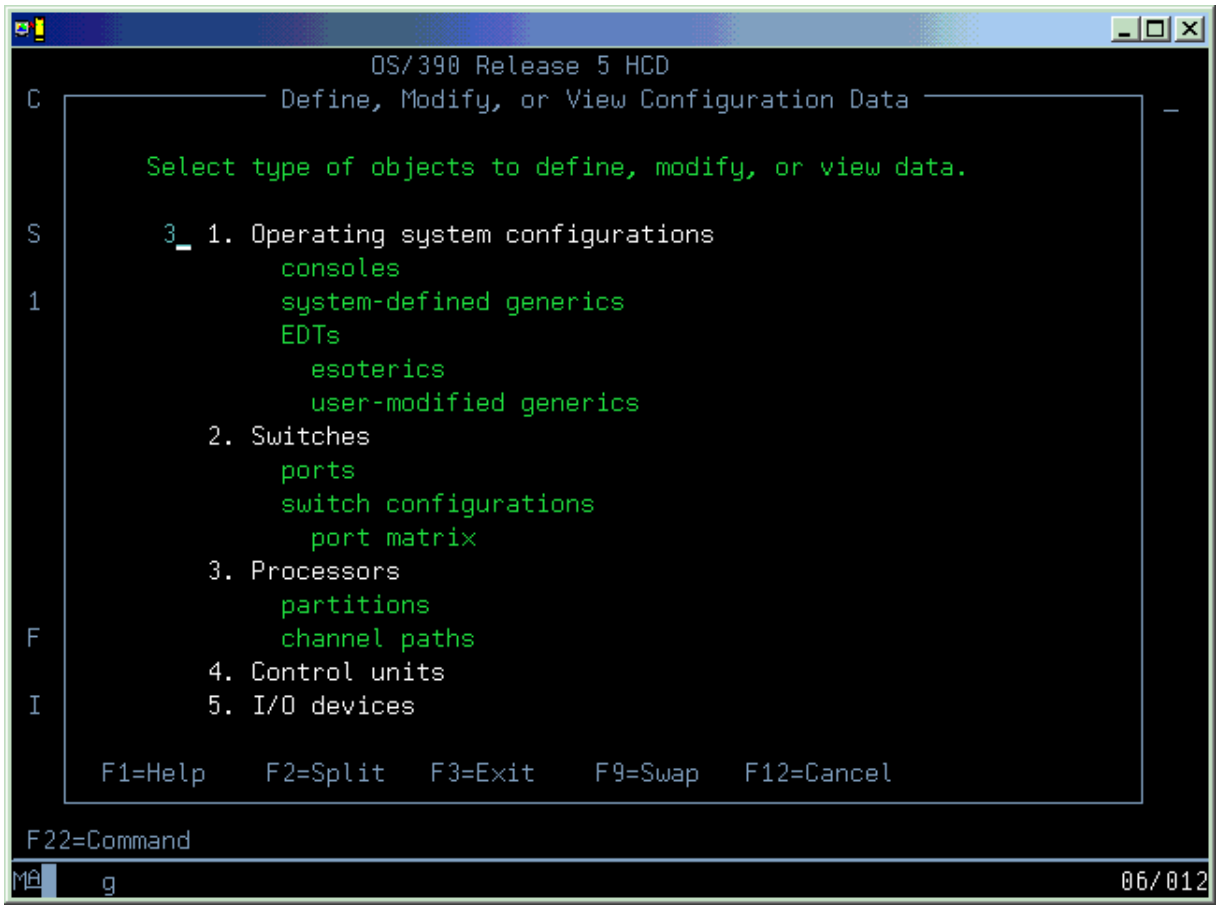


Figure 161. Select processors

2. Select 'S' Work with attached channel paths

© Copyright IBM Corp. 2000, 2022

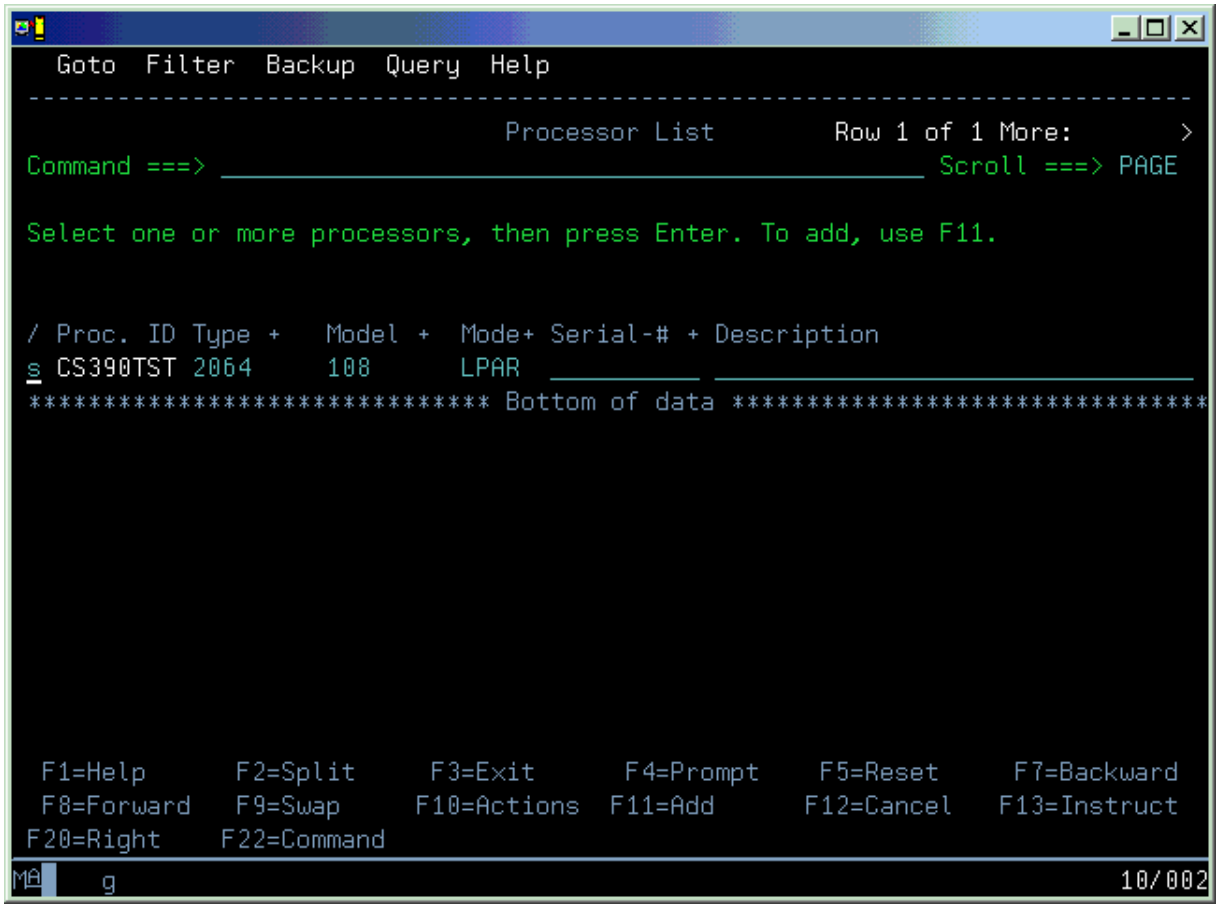


Figure 162. Work with attached channel paths

3. On the Channel Path List enter the Add command (or press F11) to initiate the Define Channel Path dialog.

Figure 163. Initiate the Define Channel Path dialog

4. Fill in the Add Channel Path panel, then press Enter (Select SHR for Operation mode to share IQD Chpids across LPARs).

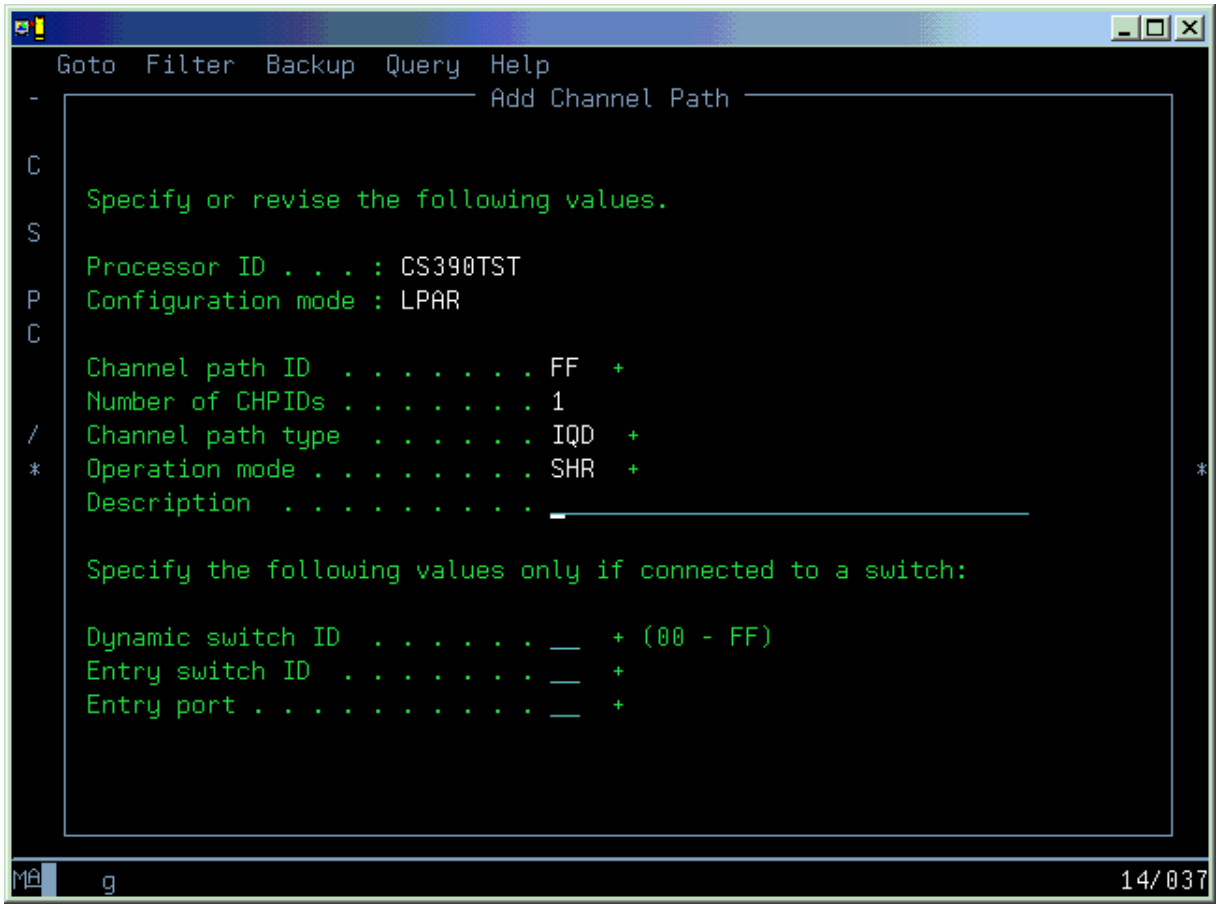


Figure 164. Add channel path

5. For an IQD channel path, the Specify Maximum Frame Size panel pops up with the default value of 16 KB.

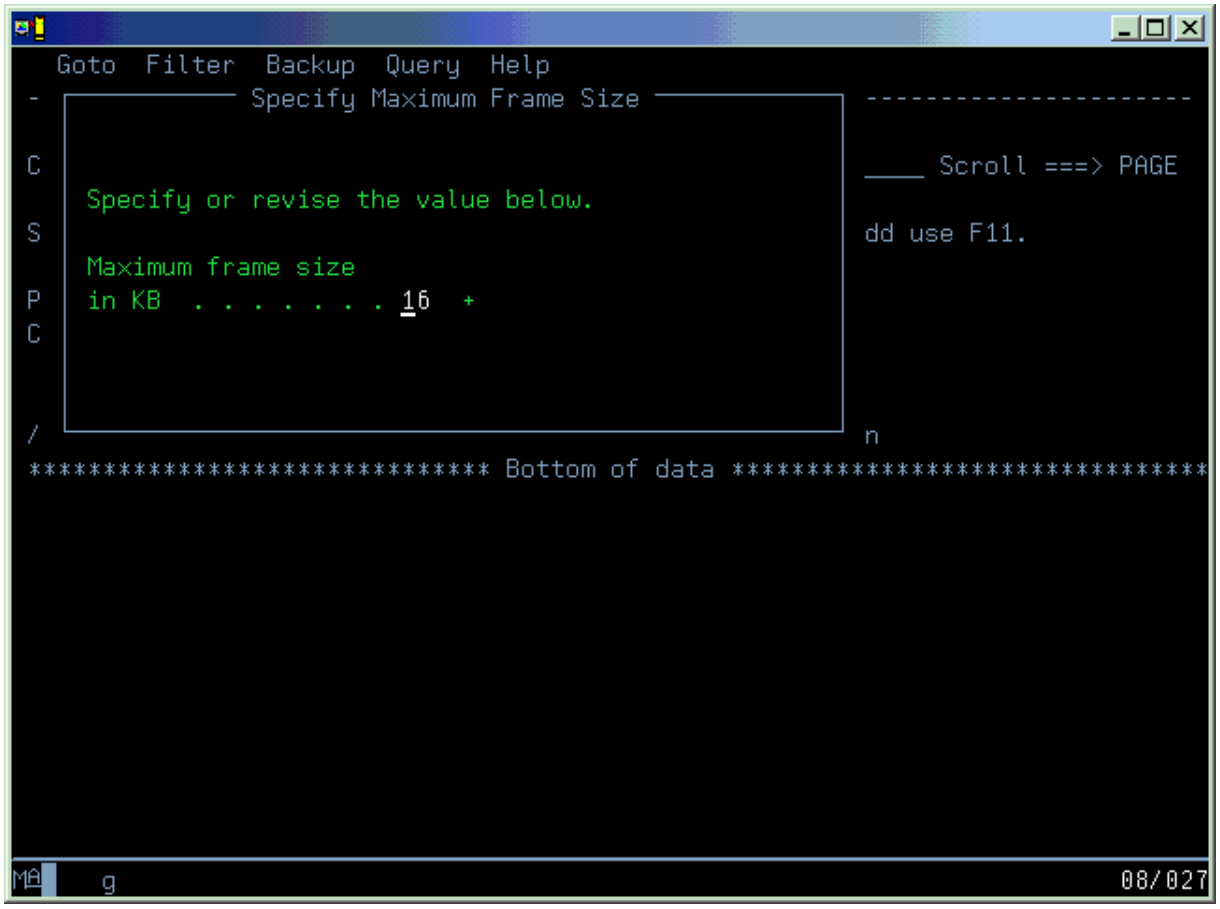


Figure 165. Specify Maximum Frame Size

Or change the frame size to wanted size:

Table 74. Frame size specification

Maximum Frame Size	TCP/IP MTU size
16K	8K
24K	16K
40K	32K
64K	56K

6. Define the channel path access list that each LPAR should have access to.

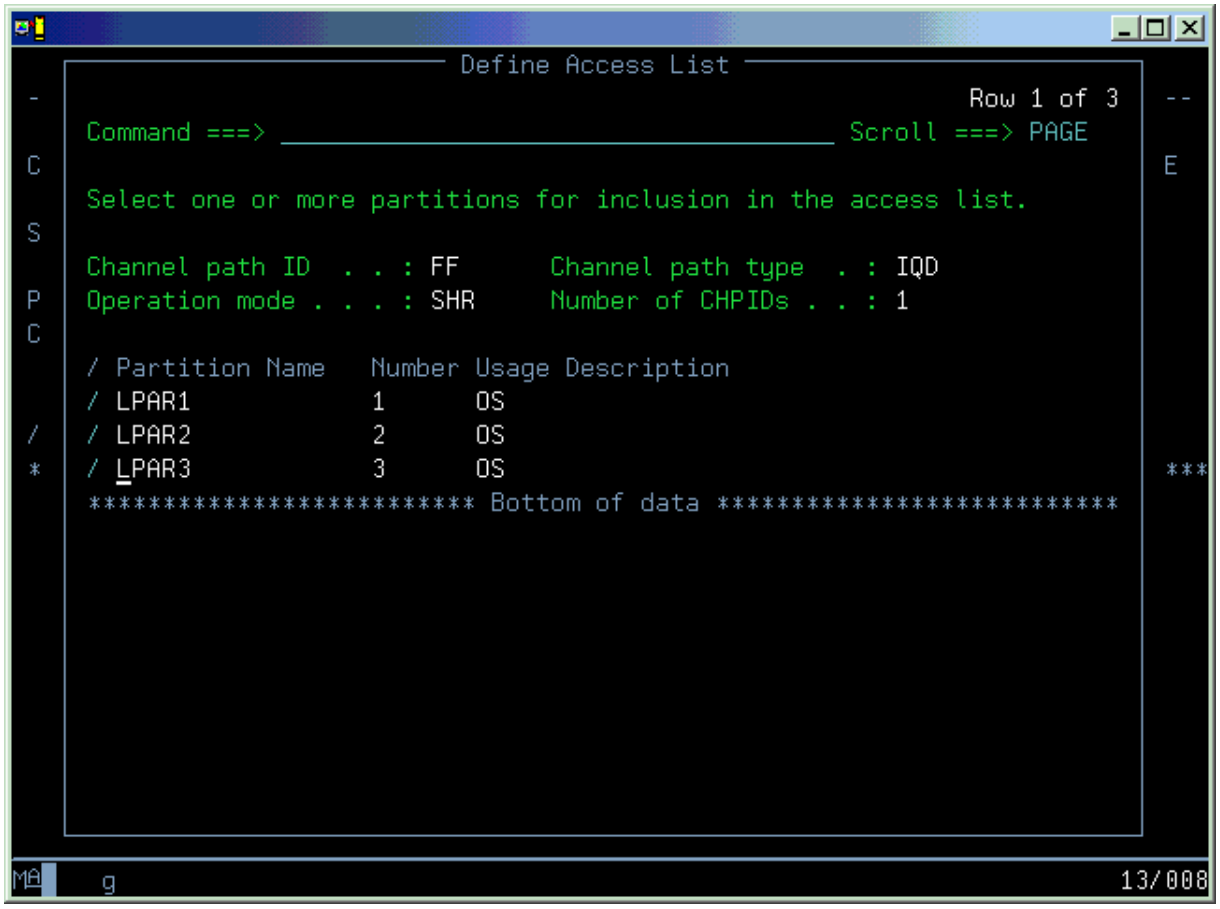


Figure 166. Define the channel path access list

7. Having pressed Enter, the Channel Path List is redisplayed with channel path number FF defined.

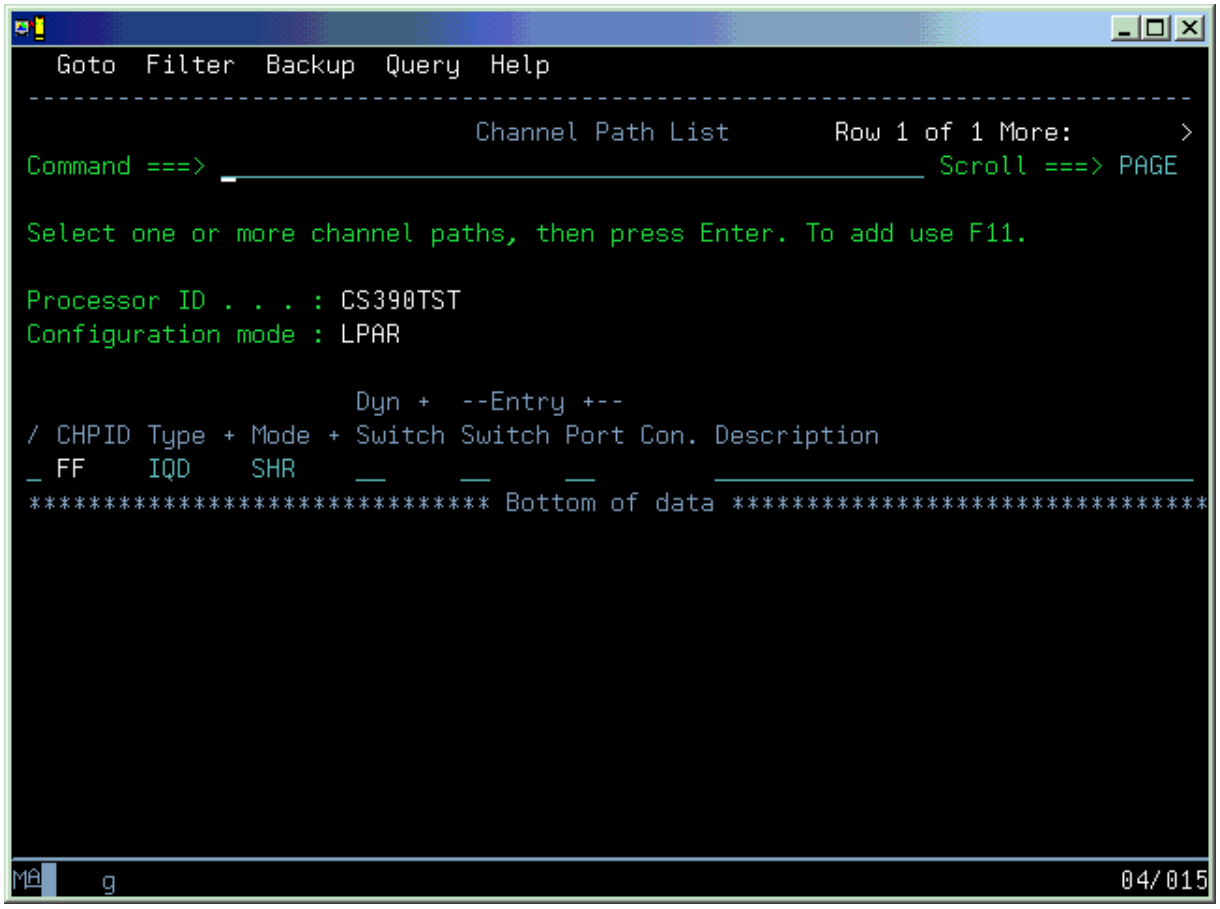


Figure 167. Channel path number FF defined

8. As the next step, add the control unit(s) to the IQD channel path. Select the defined channel path with action "Work with attached control units" (action code 'S').

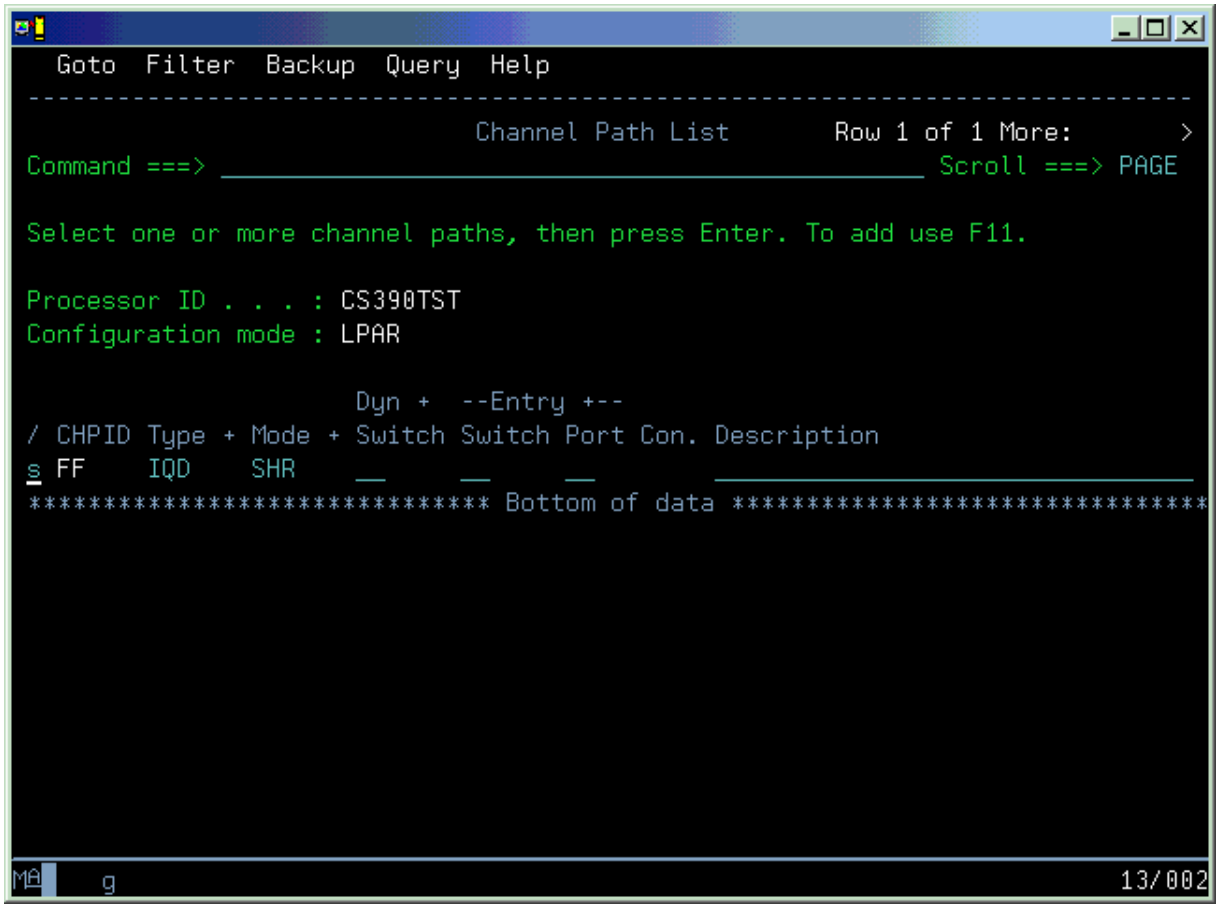


Figure 168. Work with attached control units

9. An empty control unit list is displayed. Enter the 'Add' command or F11.

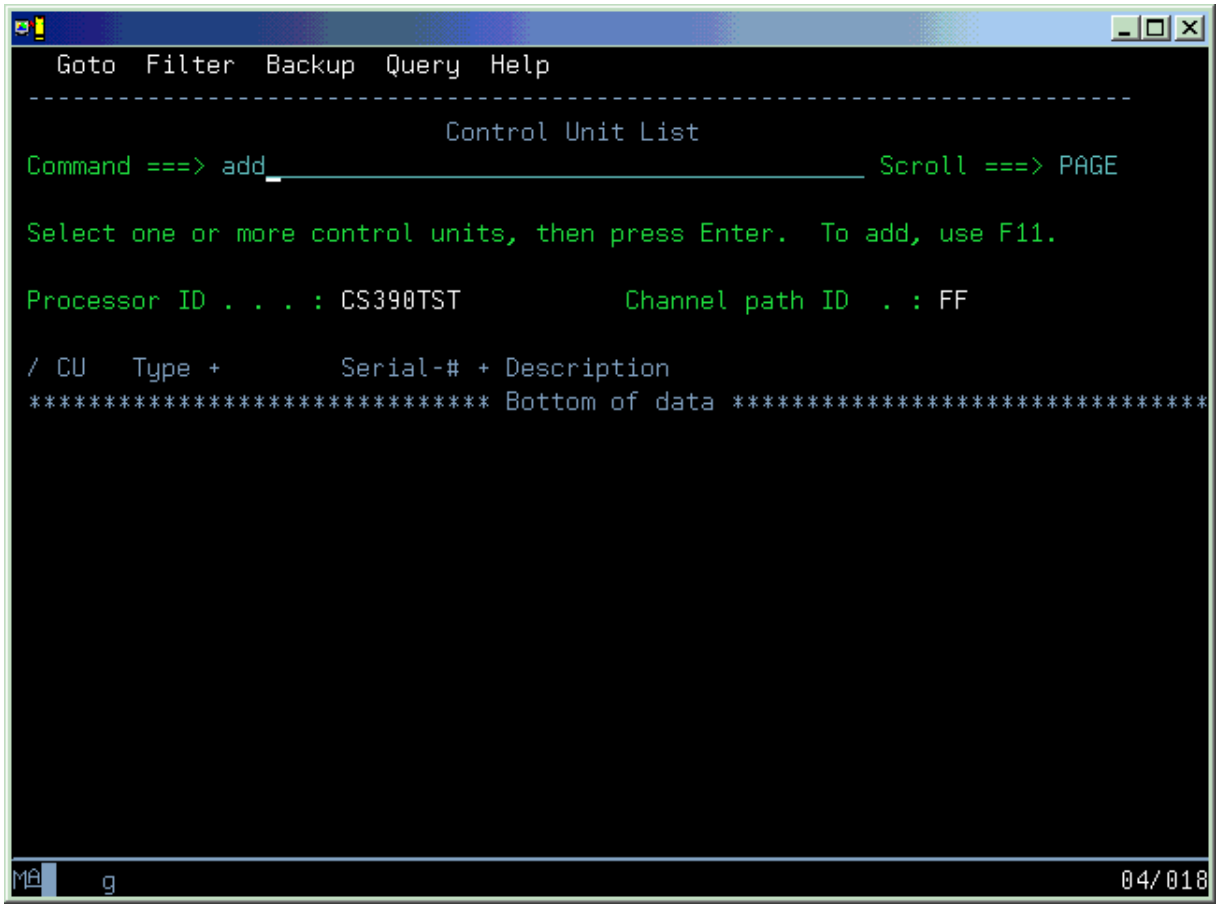


Figure 169. Add the control unit(s)

10. Define a control unit of type 'IQD' for channel path FF.

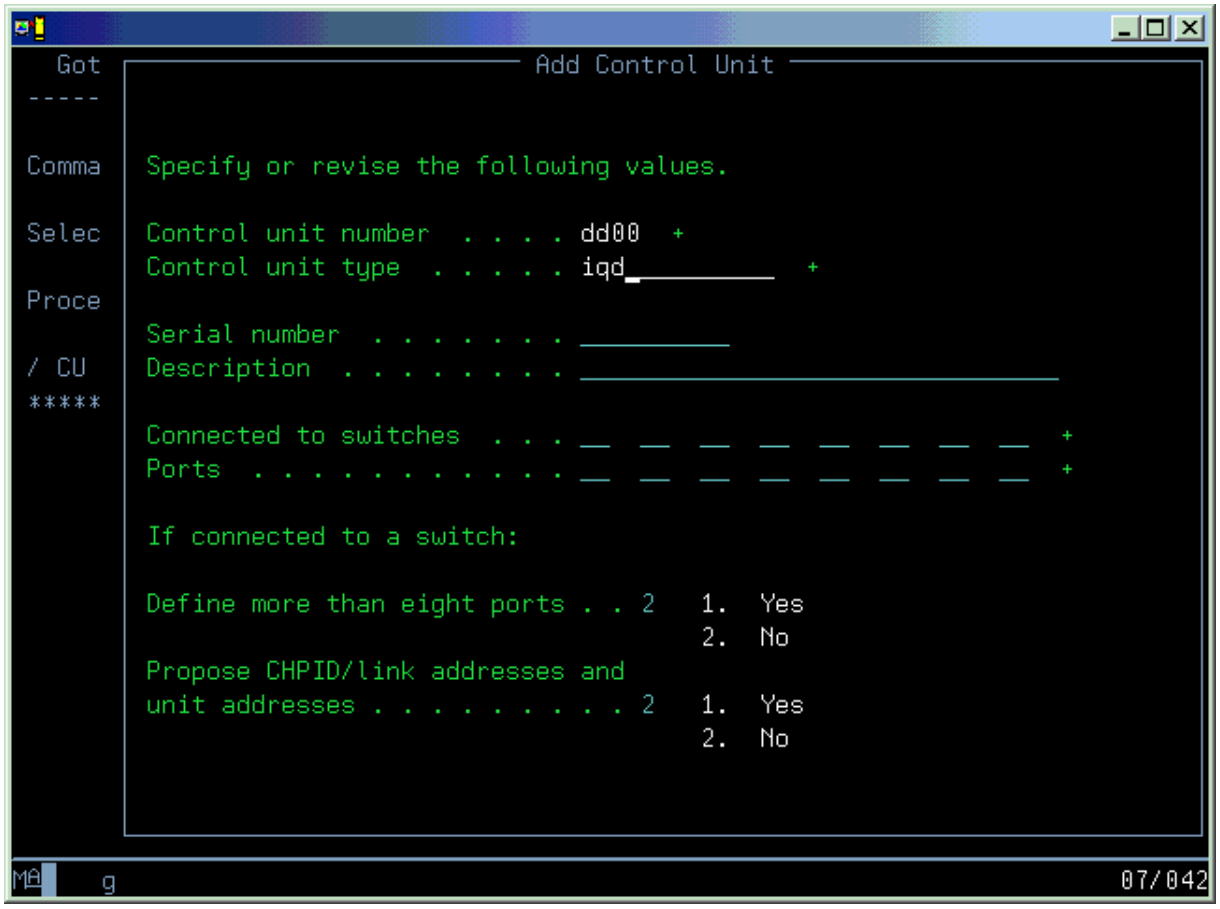


Figure 170. Define a control unit

11. Define it to the processor:

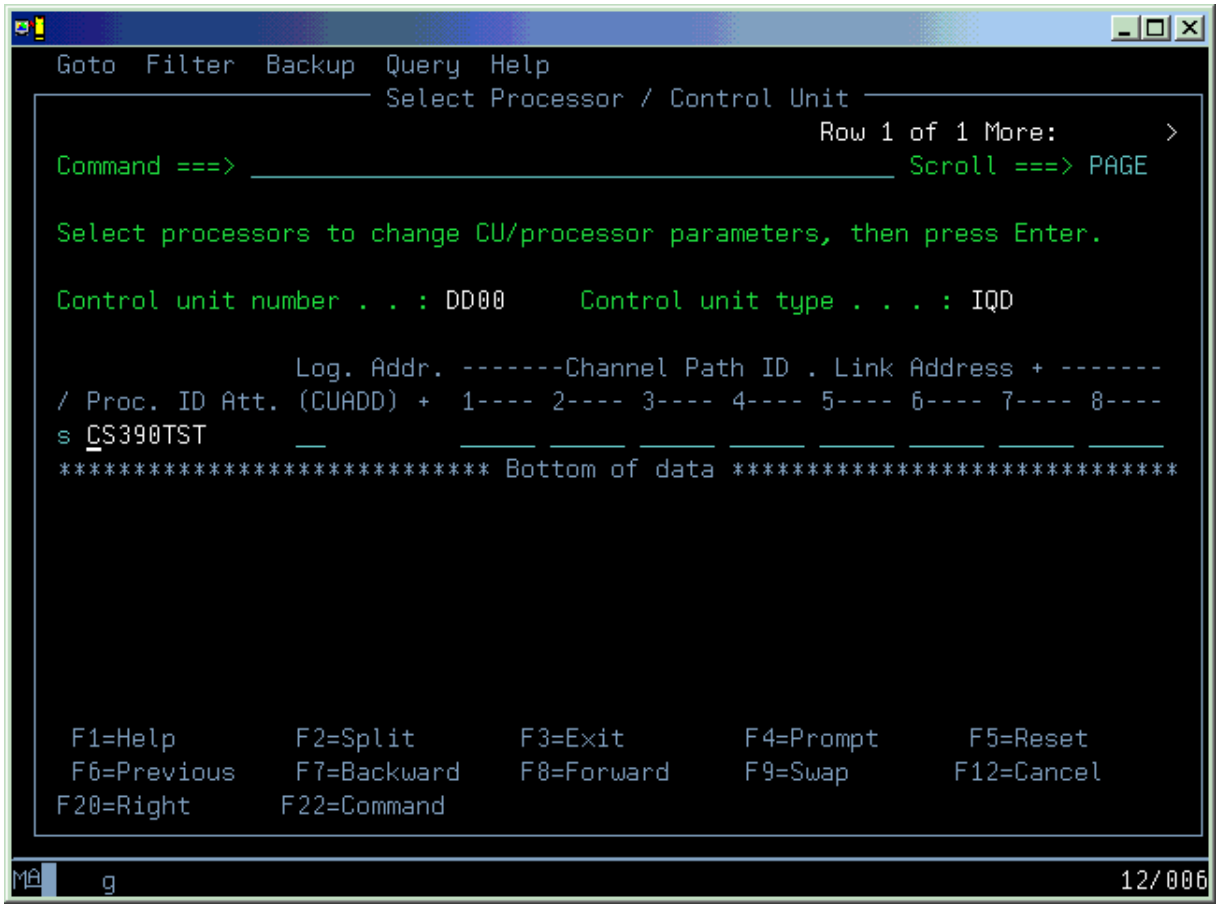


Figure 171. Define it to the processor

12. The processor settings are already preset. Pressing Enter, returns to the Select Processor/Control unit panel. Pressing Enter again, returns to the Control Unit List panel which shows the currently defined control unit.

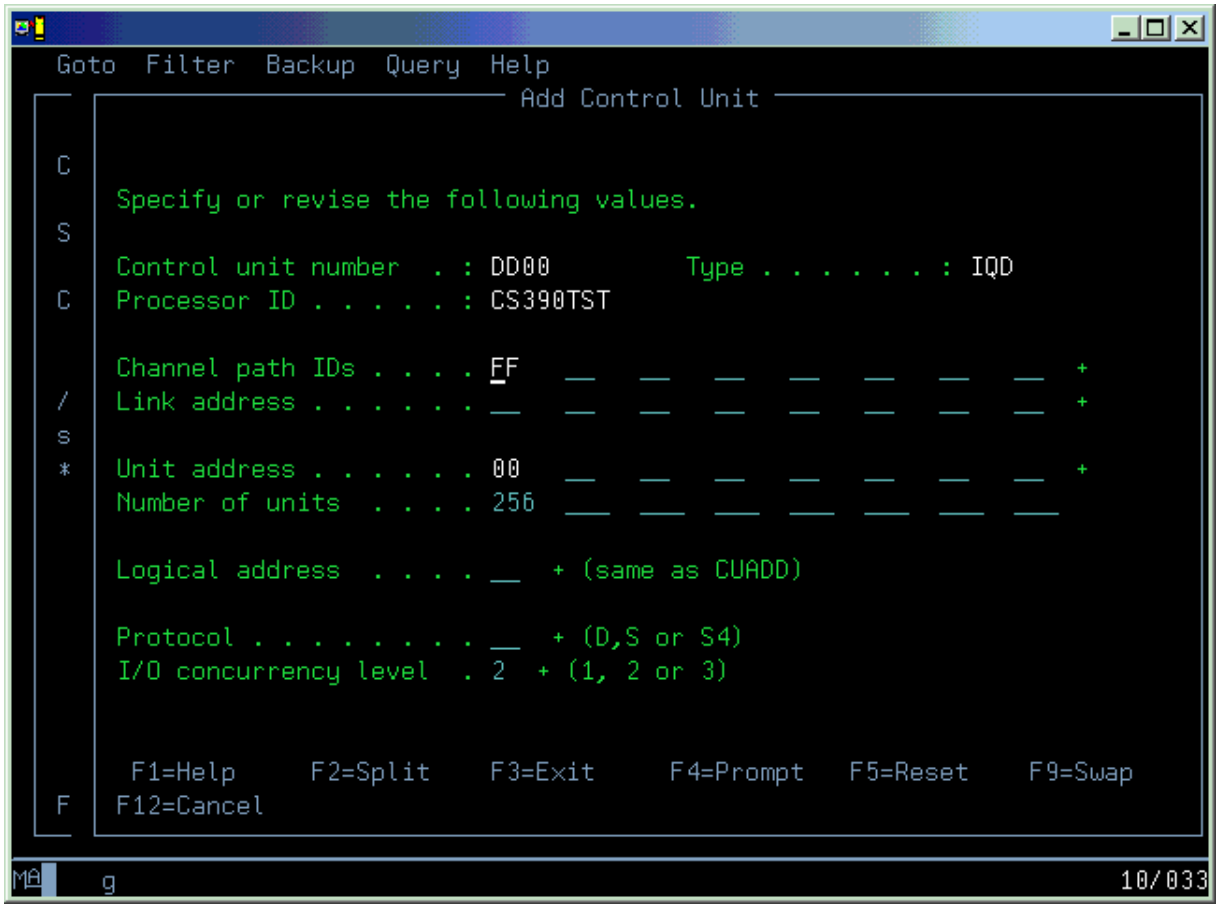


Figure 172. Currently defined control unit

13. Next, define the devices. Successively, select a control unit and perform action "Work with attached Devices".

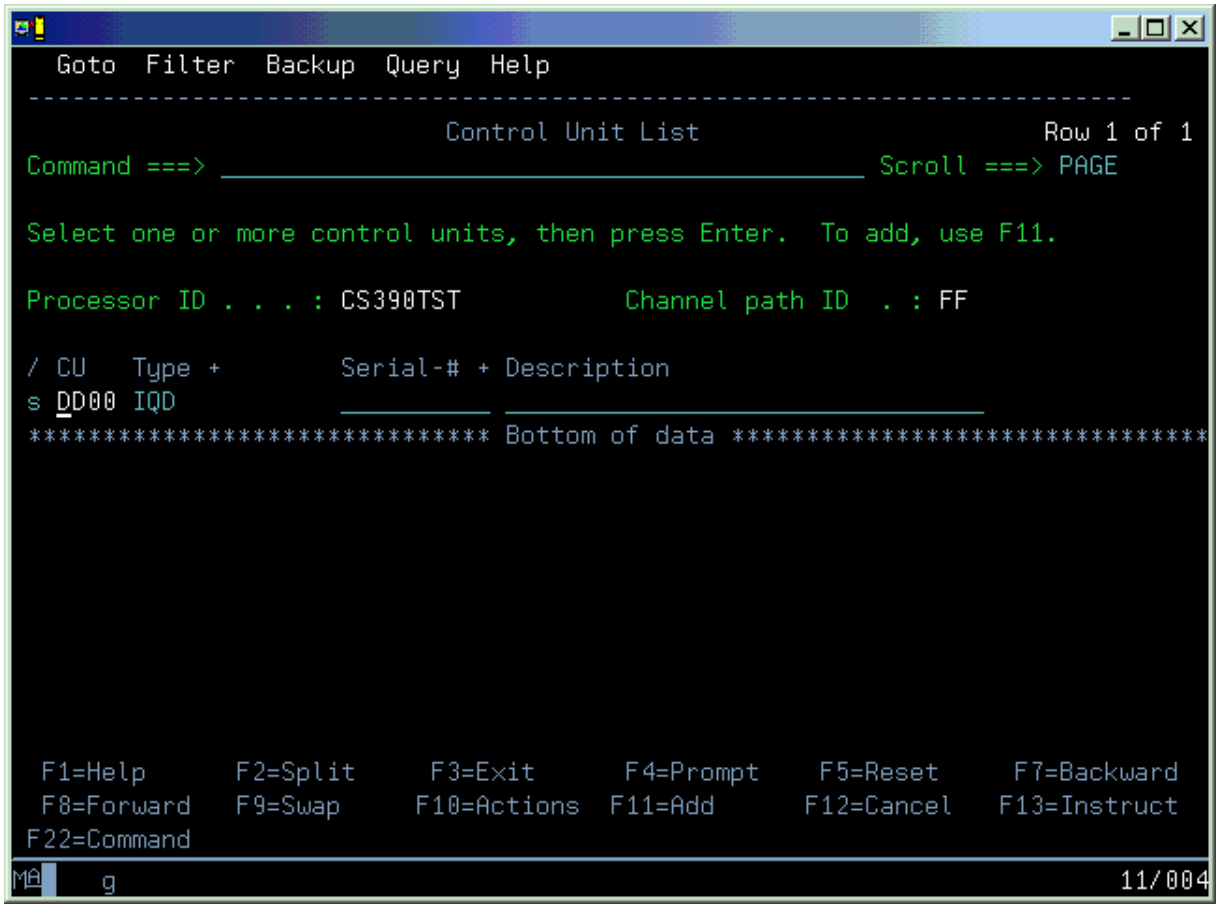


Figure 173. Define the devices

14. This leads to an empty device list.

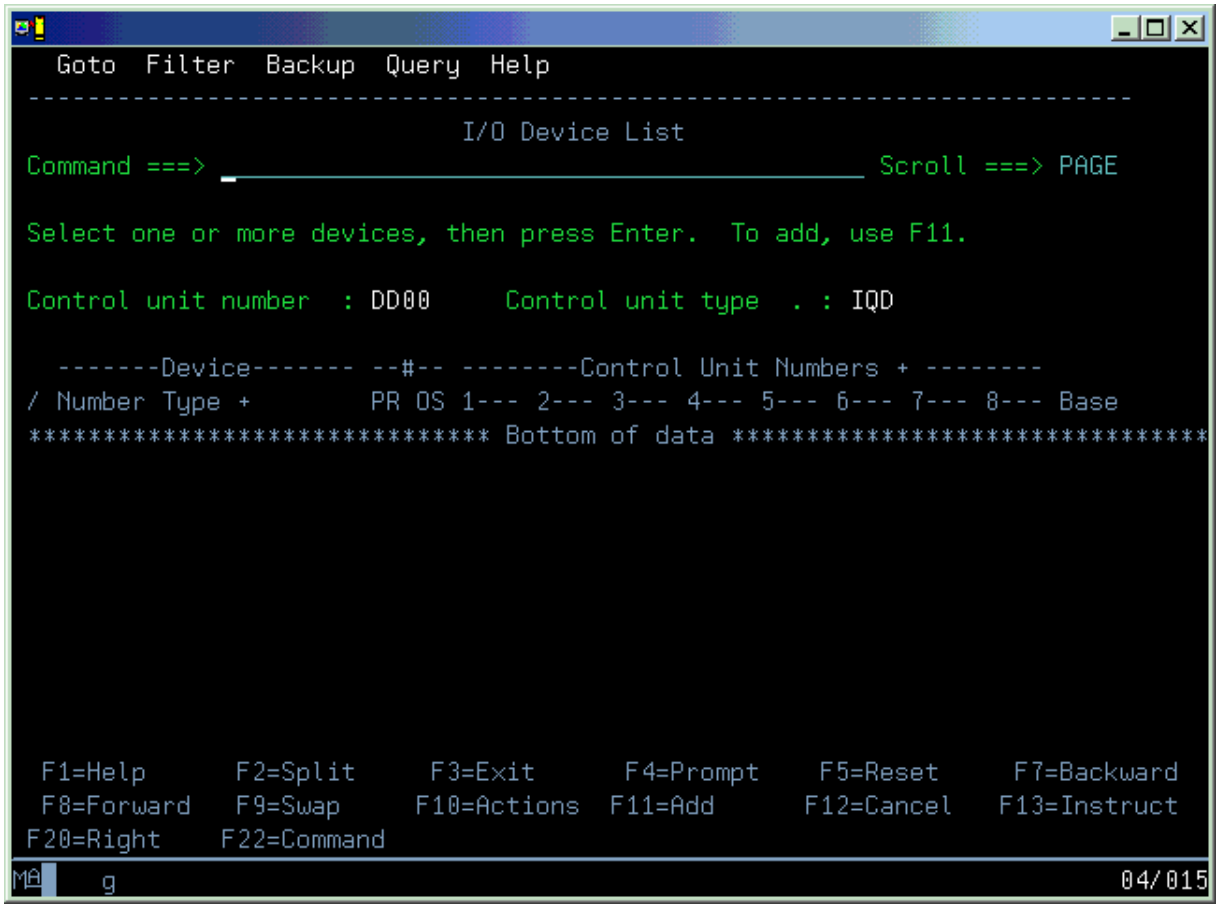


Figure 174. Empty device list

15. Perform the Add action to define the devices for the control unit selected in the previous step.

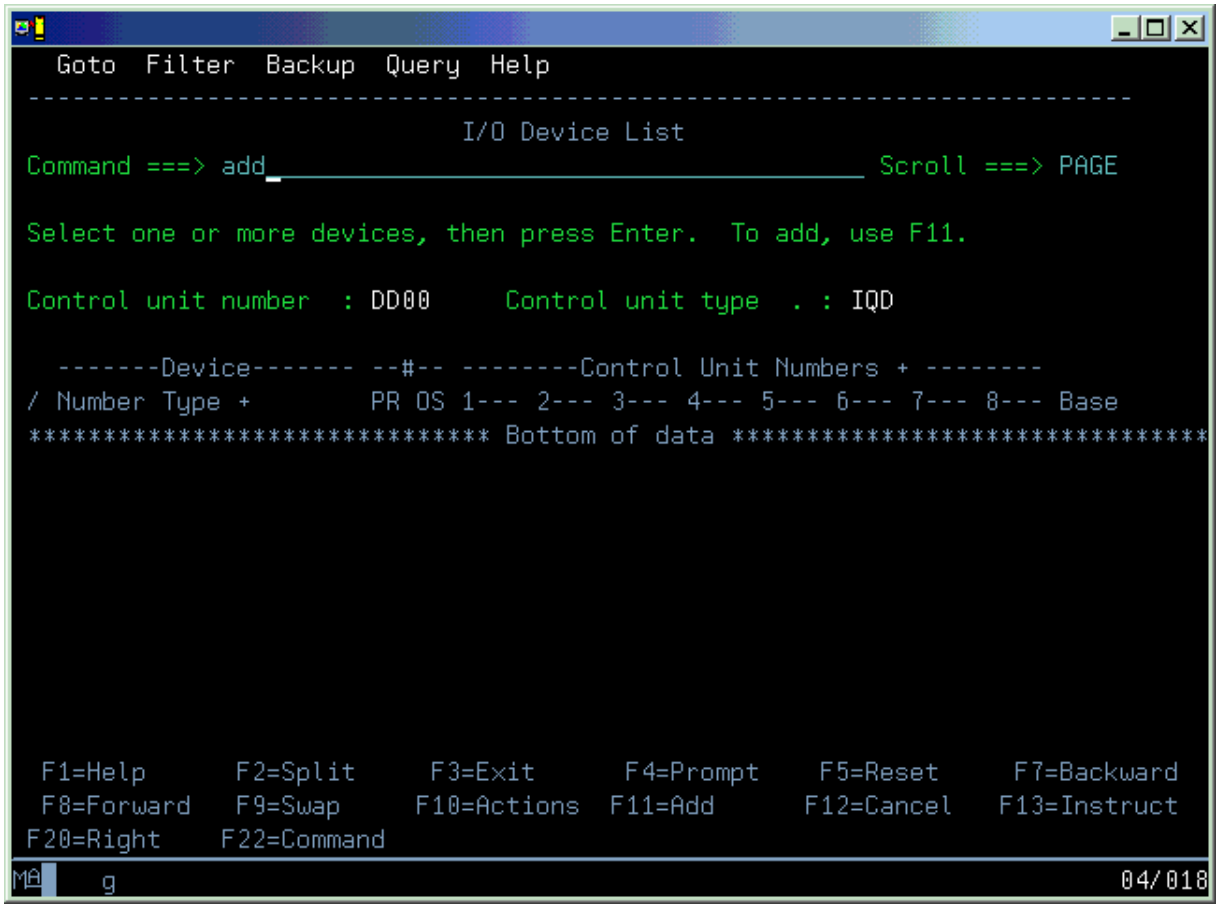


Figure 175. Define the devices for the control unit

16. Add devices of type IQD to the selected control unit.

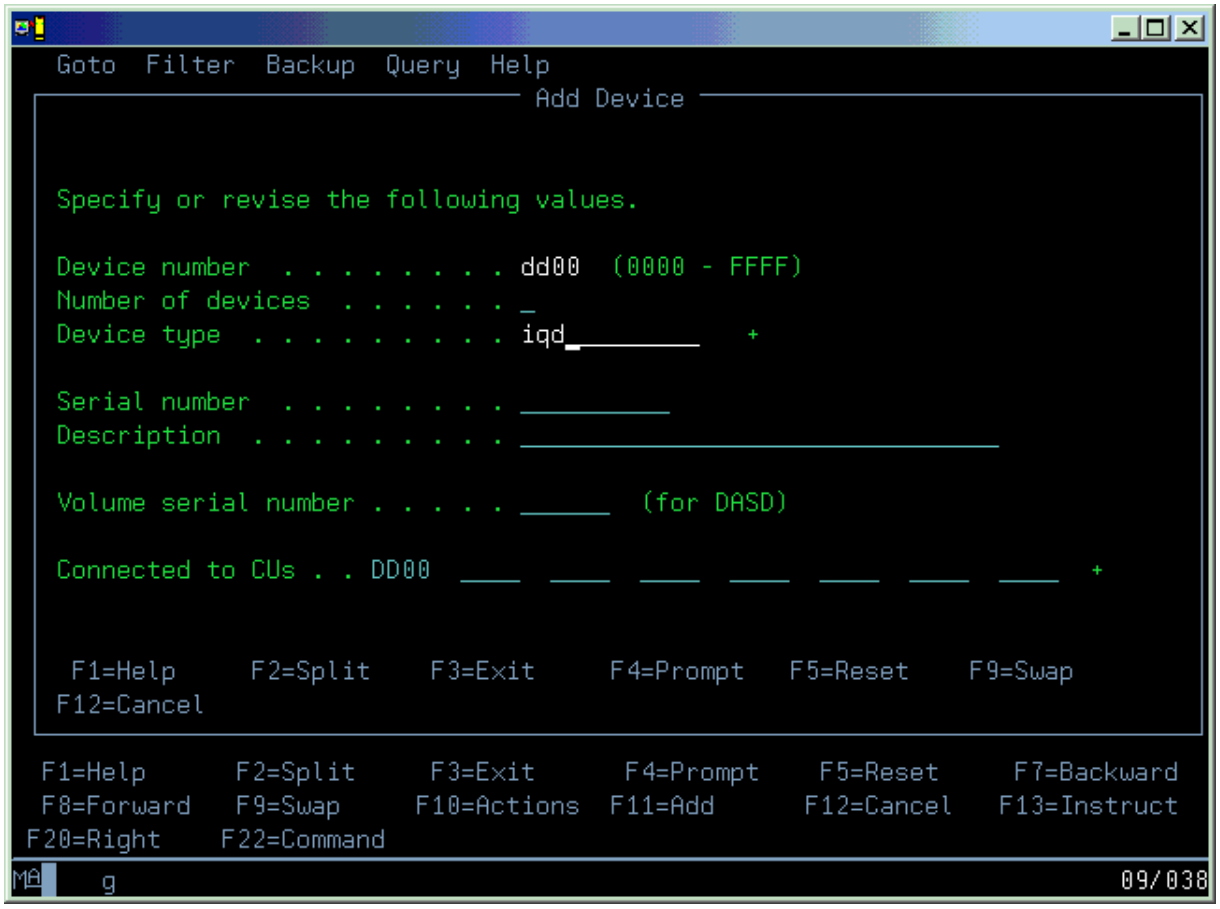


Figure 176. Add devices of type IQD

17. If the number of devices has been left unspecified (as in this example), HCD defines 10 devices.

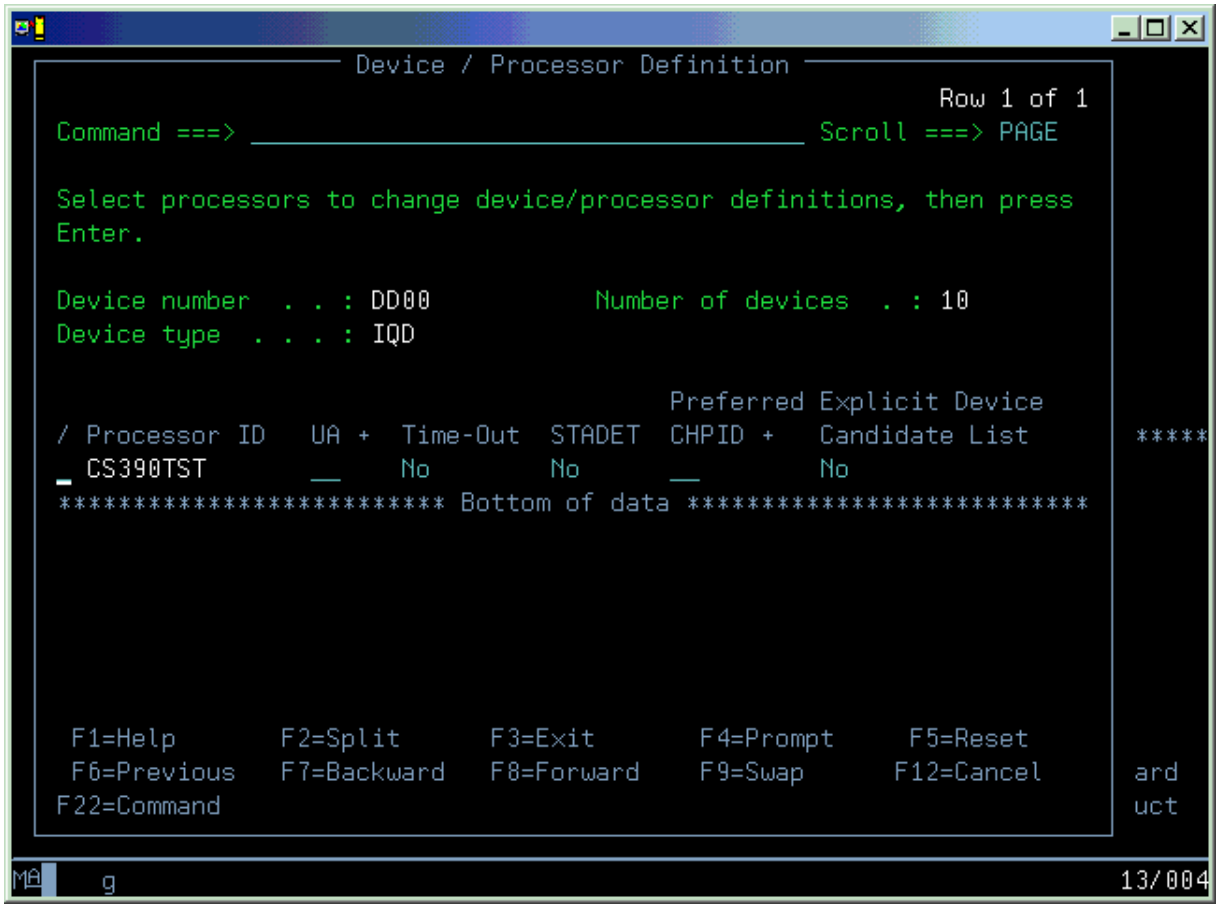


Figure 177. Define number of devices

18. Hit enter and the next panel displayed will be:

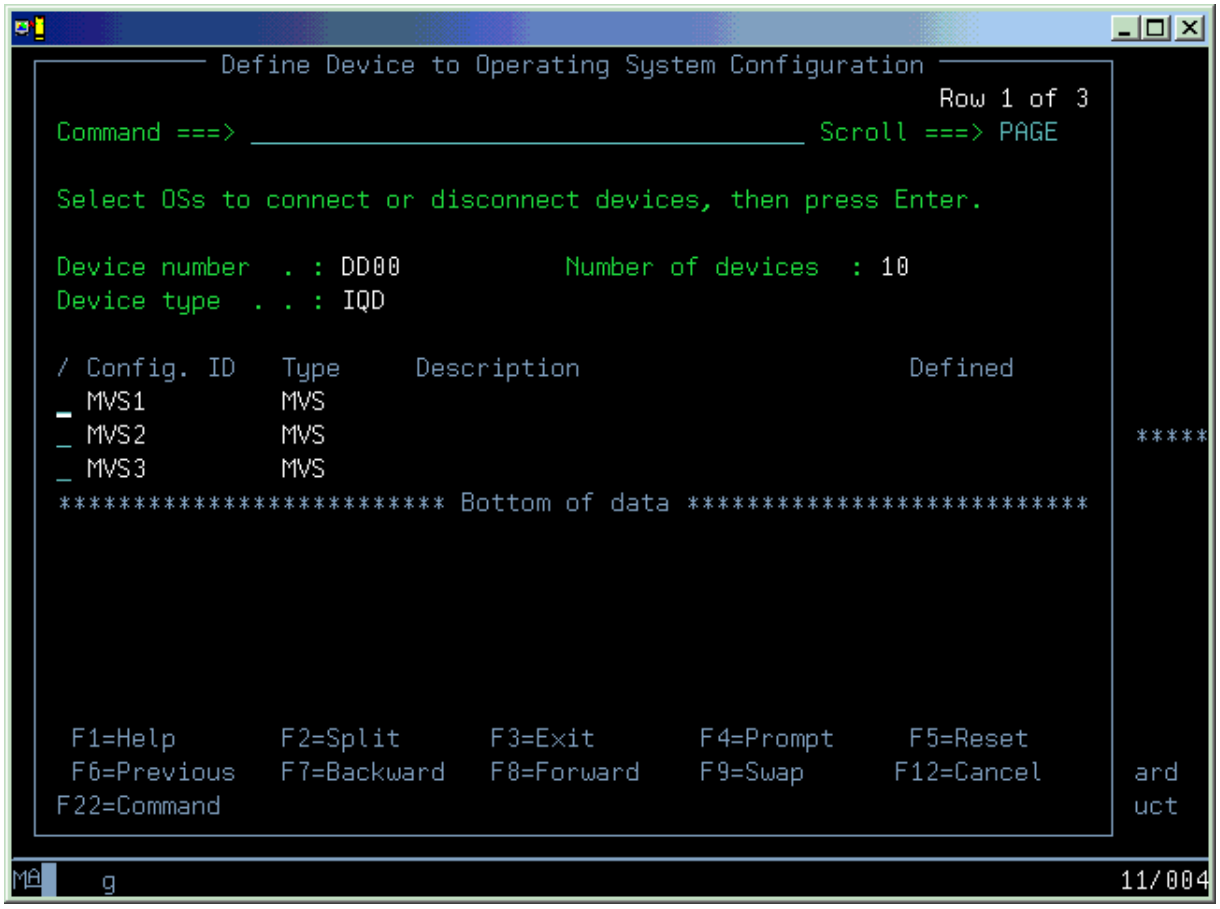


Figure 178. Define device to operating system

19. Next, define the devices to the operating system by selecting an 'S' on each system you want to have them defined on.

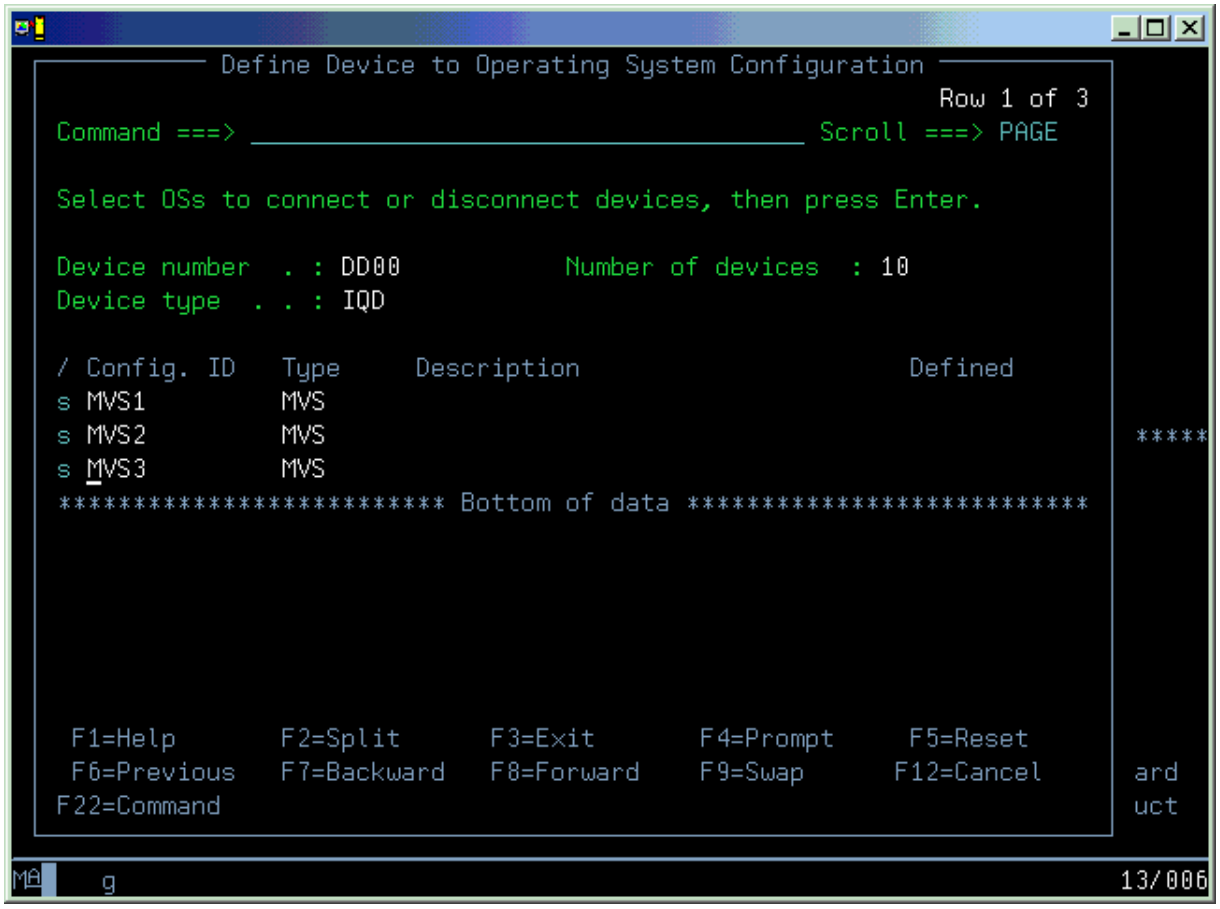


Figure 179. Select systems

20. The device parameters are shown with default values. Press Enter, to complete the definition for each system.

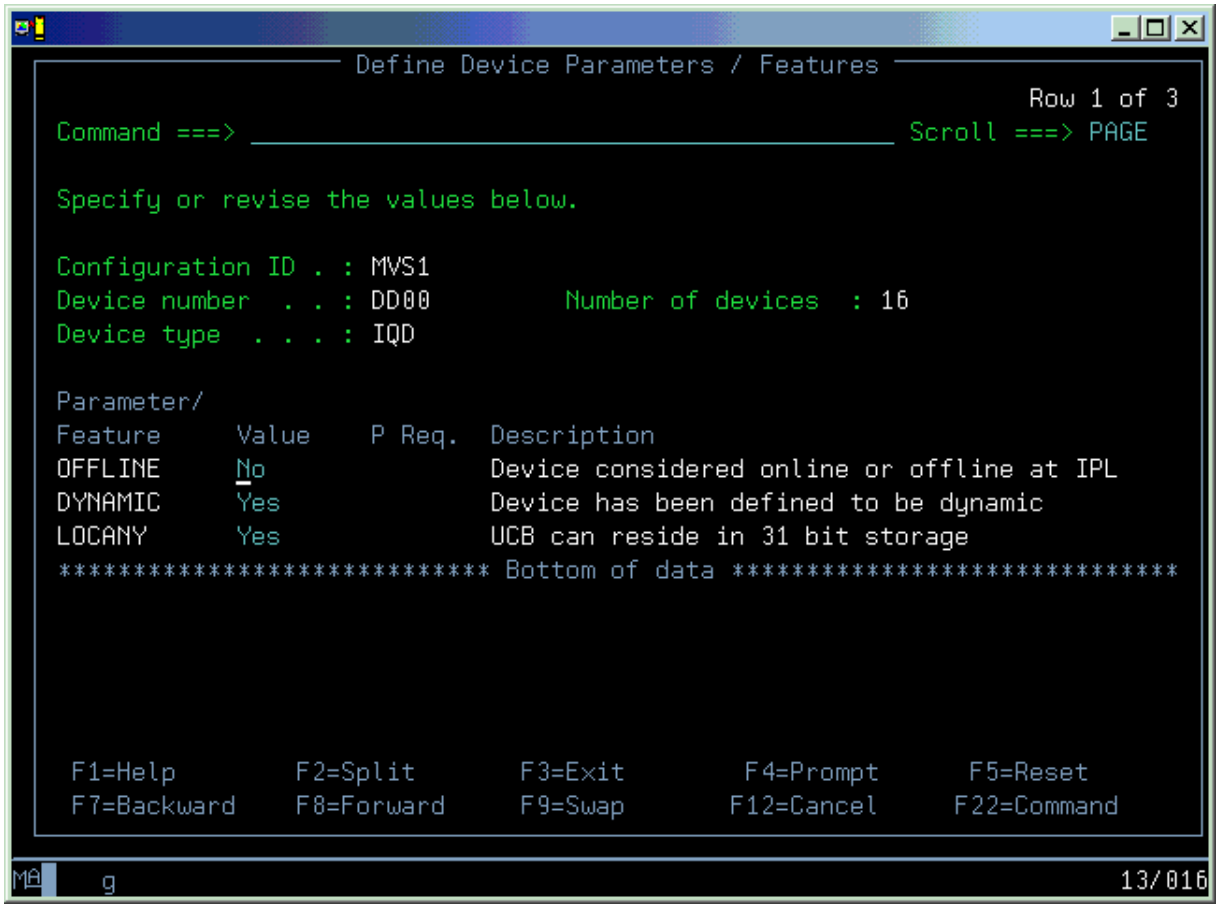


Figure 180. Complete the definition

21. Press Enter until you return to the I/O device list panel, the definition for channel path FF is complete.

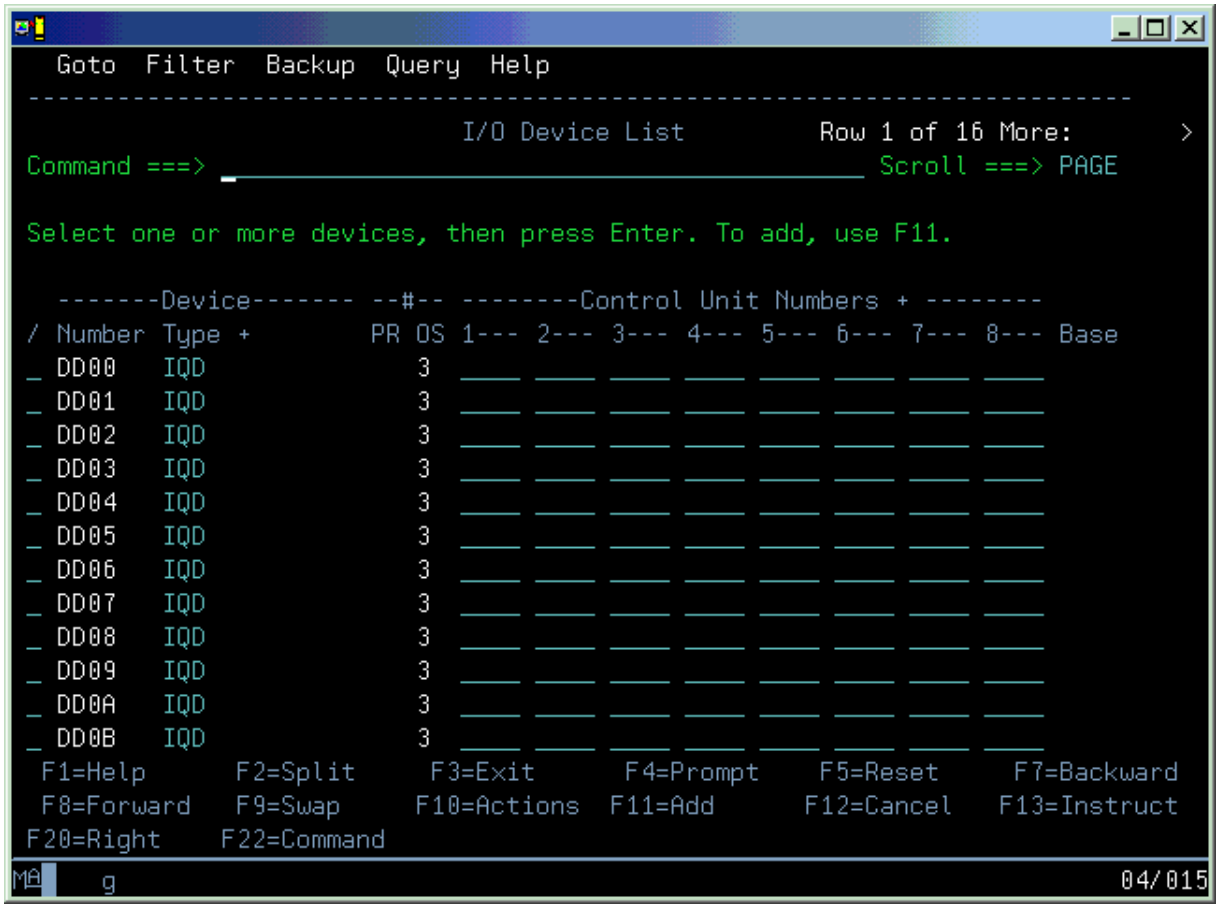


Figure 181. Definition completed

The sample IOCP input for this example would be:

```
----+----1----+----2----+----3----+----4----+----5----+----6----+----7--
         CHPID PATH=(FF),SHARED,                                       *
               PARTITION=((LPAR1,LPAR2,LPAR3),(LPAR1,LPAR2,LPAR3)),    *
               TYPE=IQD,OS=00
         CNTLUNIT CUNUMBR=DD00,PATH=(FF),UNIT=IQD
         IODEVICE ADDRESS=(DD00,010),CUNUMBR=(DD00),UNIT=IQD
```

Appendix E. Steps for preparing to run IP security

To run IP security, you need to configure the IKE daemon. Each step is described in detail in the corresponding subtopic.

Procedure

Perform the following steps to prepare to run the IKE daemon:

1. Set the appropriate UNIX System Services parameters.
2. Ensure the IKE daemon's user ID has enough system resources.
3. Authorize the IKE daemon to the external security manager.
4. Authorize the **ipsec** command to the external security manager.
5. Authorize IP security to ICSF (optional).
6. Set up the IKE daemon for RSA signature mode authentication (optional).

Step 1: Setting appropriate UNIX System Services parameters

Verify that AF_UNIX and AF_INET are defined in the BPXPRMxx member of SYS1.PARMLIB. If the domains are not defined, see z/OS UNIX System Services Planning for the steps to customize the BPXPRMxx parmlib member.

Step 2: Ensure the IKE daemon's user ID has enough system resources

The user ID under which the IKE daemon runs requires sufficient system resources when all the z/OS system's IKE peers attempt to negotiate or renegotiate IPSec tunnels with z/OS at the same time. If thousands of IKE peers attempt to negotiate or renegotiate, you might need to adjust certain system parameters upwards to ensure that IKED does not run out of critical resources. Ensure that IKED does not run out of the following critical resources:

- Virtual storage
- System message queue space
- UDP queue capacity

Follow the instructions in this section to ensure that the IKE daemon's user ID has sufficient resources.

Virtual storage capacity

Take the following steps to ensure the user ID has enough virtual storage:

- Adjust the REGION parameter to ensure that the TSO/E region size (REGION) for the IKE daemon user ID is at least 1 GB.
- Set the BPXPRMxx parmlib member by using the MAXASSIZE parameter to ensure that ASSIZEMAX in the OMVS segment for your user profile is at least 1 GB.
- Set the MEMLIMIT value in your OMVS segment, or set the BPXPRMxx or the SMFPRMxx parmlib members to ensure that the MEMLIMIT value for the user ID is at least 1 GB. The default value is 2 GB.

You can see the format of the **RACF ALTUSER** command when the command is issued to assign the storage allocations for a user. The MEMLIMIT value is set in the OMVS segment for the user.

```
ALU user-name TSO(SIZE(1048576)) OMVS(ASSIZEMAX(1073741824) MEMLIMIT(1G))
```

© Copyright IBM Corp. 2000, 2022

For more information about the MEMLIMIT system parameter, see z/OS MVS Programming: Extended Addressability Guide.For information about the **ulimit** command, see z/OS UNIX System Services Command Reference.

System message queue capacity

The IKE daemon uses internal OMVS message queues to route work to various threads in the process. In environments with a large number of IKE peers, the number of messages on these queues might grow significantly, to the point where the default maximum number of messages allowed on a message queue is exceeded. In such environments, set the IPCMSGQMNUM parameter of the SETOMVS command to a high value to accommodate scenarios where large numbers of SAs are being negotiated simultaneously. For more information about the IPCMSGQMNUM parameter, see z/OS MVS System Commands.

UDP queue capacity

If UDPQUEUELIMIT is configured or defaulted in the TCPIP profile, a UDP queue limit is in effect for the TCP/IP stack. When large numbers of IKE peers are negotiating SAs with z/OS, the limit can be exceeded. Thus, some remote IKE peers retransmit messages to z/OS and the duration of the SA negotiations increases. In such cases, you can use the UDP Traffic Regulation function of z/OS Communications Server's Intrusion Detection Services (IDS) to increase the UDP queue limit for the IKE daemon's UDP queues. Specifically, define a UDP traffic regulation rule for ports 500 and 4500 with the queue limit value set to VERY_LONG. For more information about UDP Traffic Regulation, see Traffic regulation policies for UDP ports .

Step 3: Authorizing the IKE daemon to the external security manager

You need to authorize the IKE daemon to the external security manager.

To authorize the IKE daemon to RACF, perform the steps in “Steps for authorizing the IKE daemon to RACF” on page 1362.

Steps for authorizing the IKE daemon to RACF

The commands used to authorize the IKE daemon to RACF are in EZARACF in the SEZAINST data set.

Procedure

Perform the following steps to authorize the IKE daemon to RACF:

1. Add user ID IKED, and add IKED to the STARTED class.
   - If IKED is defined using UID 0:

```
ADDUSER  IKED     DFLTGRP(OMVSGRP) NOPASSWORD OMVS(UID(0)  HOME('/'))
RDEFINE  STARTED  IKED.*            STDATA(USER(IKED))
PERMIT   BPX.DAEMON CLASS(FACILITY) ID(IKED)     ACCESS(READ)
SETROPTS RACLIST(STARTED) REFRESH
SETROPTS GENERIC(STARTED)  REFRESH
```

   - If IKED is defined using a nonzero UID (for example, 300) and IKE GID (for example, 931), then IKED needs the following access:

```
ADDGROUP IKE OMVS(GID(931))
ADDUSER  IKED DFLTGRP(IKE) OMVS(UID(300) HOME('/var/ike/'))  NOPASSWORD
CONNECT  IKED GROUP(IKE) UACC(READ)
RDEFINE  STARTED  IKED.*            STDATA(USER(IKED))
PERMIT   BPX.DAEMON CLASS(FACILITY) ID(IKED)     ACCESS(READ)
SETROPTS RACLIST(STARTED) REFRESH
SETROPTS GENERIC(STARTED) REFRESH
```

     a. Set the /var directory access to all using the following command:

```
chmod 777 /var
```

b. If the /var/ike directory does not already exist, use a superuser ID to create it and modify this directory using the following commands:

```
mkdir /var/ike/
chown IKED /var/ike
chgrp IKE /var/ike
chmod 770 /var/ike
```

c. If the /var/sock/ directory does not already exist, use a superuser ID to create it and set the access to all using the following commands:

```
mkdir /var/sock/
chown 0 /var/sock
chmod 777 /var/sock
```

d. Define RACF definitions for IKED to retrieve IP security policies from Policy Agent:

```
RDEFINE SERVAUTH EZB.PAGENT.sysname.*.IPSEC UACC(NONE)
PERMIT EZB.PAGENT.sysname.*.IPSEC CLASS(SERVAUTH) ID(IKED) ACCESS(READ)
SETROPTS RACLIST(SERVAUTH) REFRESH
SETROPTS GENERIC(STARTED)  REFRESH
```

e. Permit IKED to issue console messages directly without the BPXM023I message prefix:

```
RDEFINE FACILITY BPX.CONSOLE UACC(NONE)
PERMIT BPX.CONSOLE CLASS(FACILITY) ID(IKED) ACCESS(READ)
SETROPTS RACLIST(FACILITY) REFRESH
```

2. Allow the IKED to access SYS1.PARMLIB as follows:

```
PERMIT   SYS1.PARMLIB  ID(IKED)        ACCESS(READ)
```

3. Enable the IKED to run as nonswappable.

   If you have defined the BPX.STOR.SWAP resource to RACF, you can enable the IKED using the following commands:

```
PERMIT BPX.STOR.SWAP CLASS(FACILITY) ID(IKED) ACCESS(READ)
SETROPTS RACLIST(FACILITY) REFRESH
```

Step 4: Authorizing the ipsec command to the external security manager

You need to authorize the **ipsec** command to the external security manager.

To authorize the **ipsec** command to RACF, perform the steps in "Steps for authorizing the ipsec command to RACF" on page 1363.

For more information about **ipsec** command security and the SERVAUTH profile, see z/OS Communications Server: IP System Administrator's Commands.

Steps for authorizing the ipsec command to RACF

The commands used to authorize the **ipsec** command are in EZARACF in the SEZAINST data set.

Procedure

Perform the following steps to authorize the **ipsec** command to RACF:

1. Define access control for the **ipsec** command.

   The **ipsec** command uses both display and control features. You can control access to each feature independently.

- To control access to both the display and control capabilities of the **ipsec** command, issue the following commands:

```
SETROPTS GENERIC(SERVAUTH)
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.tcpprocname.* UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.tcpprocname.* CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
SETROPTS GENERIC(SERVAUTH) REFRESH
```

- To control access specifically to the display capabilities of the **ipsec** command for a stack, issue the following commands:

```
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.tcpprocname.DISPLAY UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.tcpprocname.DISPLAY CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
```

- To control access specifically to the display capabilities of the **ipsec** command for global defensive filters, issue the following commands:

```
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.DMD_GLOBAL.DISPLAY UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.DMD_GLOBAL.DISPLAY CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
```

- To control access specifically to the control capabilities of the **ipsec** command for a stack, issue the following commands:

```
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.tcpprocname.CONTROL UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.tcpprocname.CONTROL CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
```

- To control access specifically to the control capabilities of the **ipsec** command for global defensive filters, issue the following commands:

```
RDEFINE  SERVAUTH EZB.IPSECCMD.sysname.DMD_GLOBAL.CONTROL UACC(NONE)
PERMIT   EZB.IPSECCMD.sysname.DMD_GLOBAL.CONTROL CLASS(SERVAUTH) ID(userid)
ACCESS(READ)
```

Tip: These SERVAUTH profiles provide **ipsec** command access to only the local stack. For information about SERVAUTH profiles for controlling **ipsec** command access for the network security services (NSS) server, see "Network security services for the IPSec discipline" on page 185.

2. To refresh the in-storage RACF profiles in the SERVAUTH class, issue the following command:

```
SETROPTS RACLIST(SERVAUTH) REFRESH
```

Step 5: Authorizing IP security to ICSF (optional)

IP security can take advantage of the encryption and decryption functions that are available on System z hardware in the following ways:

- Encrypting and decrypting TCP/IP packet data in an IPSec tunnel.
- Encrypting signature data included as part of IKE message flows. This encryption is performed only when digital signature authentication is requested.

This encryption support is provided by the combination of the Integrated Cryptographic Feature (ICRF) on the processor and the Integrated Cryptographic Service Facility/MVS (ICSF) software product. ICSF provides cryptography support through various cryptographic hardware features. The cryptographic features that are available to your applications depends on your processor or server model. For information about which features are available on your hardware, see the information about callable service support by hardware configuration in z/OS Cryptographic Services ICSF Overview.

To use this support, ICSF must be started and running. Preferably, start ICSF prior to starting TCP/IP. However, it can also be started when TCP/IP is active. For details on configuring ICSF, see z/OS Cryptographic Services ICSF Administrator's Guide. ICSF provides SAF controls that you can optionally

use to restrict access to these cryptographic services. To view a sample procedure for generating the corresponding SAF profiles for various CSFSERV services, see the `Cryptographic Services Authorization` section of the EZARACF sample in the SEZAINST data set.

Requirement: If you plan to control access to the ICSF cryptographic support, TCP/IP must be permitted to access the ICSF cryptographic services (CSFSERV).

Guideline: If you do not have any reason to restrict access to the ICSF cryptographic support, you should not activate the CSFSERV resource class, define any of the profiles listed below, or permit users to these profiles. If you do need to set up controls in the CSFSERV resource class, complete the steps below to enable use of ICSF for IP security.

Steps for setting up profiles in the CSFSERV resource class

Use these steps to control access to the ICSF cryptographic support.

Procedure

Perform the following steps to set up profiles in the CSFSERV resource class:

1. Determine the SAF profiles that you will use within the CSFSERV resource class:
   a) You must permit the IKED and the NSSD to the CSFIQF profile.
   b) If you have CP Assist for Cryptographic Function (CPACF) enabled on your processors and you want to take advantage of it, you must have ICSF started but you do not need to grant permission to any SAF profiles.
   c) If you do not have CPACF enabled, but you do have a cryptographic coprocessor, to take advantage of it you must perform the following steps:
      i) Permit TCP/IP to the following profiles:
         - CSFCKI
         - CSFCKM (used only for Triple DES)
         - CSFDEC1
         - CSFENC1
         - CSFOWH1
      ii) Permit the IKED and the NSSD to the following profiles:
         - CSFDSG
         - CSFDSV
         - CSFPKI
   d) If you are using AES encryption in TCP/IP for manual or dynamic tunnels, you must permit TCP/IP to the following profiles:
      - CSFDEC1
      - CSFENC1
   e) If you are using AES encryption in the IKED for dynamic tunnels, you must permit the IKED to the following profiles:
      - CSFDEC
      - CSFENC
   f) If you are using SHA2 or AES-XCBC authentication in TCP/IP, you must permit TCP/IP to the following profiles:
      - CSF1HMG
      - CSF1TRC
      - CSF1TRD

- CSFMGN1

g) If you are using SHA2 or AES-XCBC authentication in the IKED, you must permit the IKED to the following profiles:

- CSF1HMG
- CSF1TRC
- CSF1TRD
- CSFOWH

h) If you are using Diffie-Hellman groups 19, 20, or 21 in the IKED, you must permit the IKED to the following profiles:

- CSF1DVK
- CSF1GAV
- CSF1GKP
- CSF1TRC
- CSF1TRD

i) If you are using digital signature authentication in the IKED, you must permit the IKED to the following profiles:

- CSFDSG
- CSFDSV
- CSFPKI

j) If you are using digital signature authentication in the NSSD, you must permit the NSSD to the following profiles:

- CSF1HMG
- CSF1TRC
- CSF1TRD
- CSFDSG
- CSFDSV
- CSFMGN
- CSFOWH
- CSFPKI

k) If you are using elliptic curve signature authentication in the NSSD, you must permit the NSSD to the following profiles:

- CSF1GAV
- CSF1PKS
- CSF1PKV

l) If you have enabled FIPS 140 support in TCP/IP, you must permit TCP/IP to the following profiles:

- CSF1HMG
- CSF1SKD
- CSF1SKE
- CSF1TRC
- CSF1TRD
- CSFRNG

m) If you have enabled FIPS 140 support in TCP/IP and you are using manual tunnels, the z/OS Communications Server Policy Agent must be permitted to the CSF1TRC profile.

n) If you have enabled FIPS 140 support in the IKED, you must permit the IKED to the following profiles:

- CSF1DMK
- CSF1DVK
- CSF1HMG
- CSF1SKD
- CSF1SKE
- CSF1TRC
- CSF1TRD
- CSFOWH

2. Define the appropriate profiles in the CSFSERV class:

```
RDEFINE CSFSERV profile-name UACC(NONE)
```

3. Give TCP/IP access to the appropriate profiles:

```
PERMIT profile-name CLASS(CSFSERV) ID(stackname) ACCESS(READ)
```

4. Give the userids associated with IKED, NSSD, and Policy Agent access to the appropriate resource profile in the CSFSERV class:

```
PERMIT profile-name CLASS(CSFSERV) ID (userid)
```

5. Activate the CSFSERV class and refresh the in-storage RACF profiles:

```
SETROPTS CLASSACT(CSFSERV)
    SETROPTS RACLIST(CSFSERV) REFRESH
```

6. Set the MAXLEN ICSF installation option to 65535 or greater because this is the maximum TCP/IP packet size. The MAXLEN installation option for hardware cryptography determines the maximum length that can be used to encrypt and decrypt data using ICSF.

Step 6: Setting up the IKE daemon for digital signature authentication (optional)

You can configure the IKE daemon to use its own certificate service or to use the certificate service from a network security services (NSS) server. You can control whether to use the native certificate service or the NSS certificate service at the stack level. A single stack can use either the native certificate service or the NSS certificate service, but it cannot use both. If the IKED is to use an IKEv2 signature-based authentication method on behalf of a stack, that stack must be defined as a network security services (NSS) client.

By default, the IKE daemon uses the native certificate service for all stacks. Certificates for stacks configured to use the native certificate service must be on the key ring specified by the KeyRing parameter of the IkeConfig statement.

The IKE daemon native certificate service does not consult certificate revocation information when it authenticates a digital signature. If you want revocation checking, then direct the IKED to use the IPSec certificate service of an NSS server and enable revocation checking for the remote security endpoint. For information about enabling revocation checking for a remote security endpoint using the KeyExchangeAction statement, see z/OS Communications Server: IP Configuration Reference.

The IKE daemon can be directed to use an NSS server's certificate service for an individual stack by specifying the Cert option on the ServiceType parameter of an NssStackConfig statement for that stack. The NssStackConfig statement is specified in the IKE daemon configuration file. The NSS server does not have to be on the same system as the IKE daemon. The location of the NSS server is specified by the NetworkSecurityServer parameter and optionally the NetworkSecurityServerBackup parameter of the

IkeConfig statement. For more information about the IkeConfig statement, see z/OS Communications Server: IP Configuration Reference.

Certificates for stacks configured to use an NSS server for certificate service must be on the key ring specified on the KeyRing parameter of the NssConfig statement in the NSS server configuration file. For more information about the NssConfig statement, see z/OS Communications Server: IP Configuration Reference.

Figure 182 on page 1369 shows a partial configuration for the IKE daemon on system SYSTEMA and an NSS server. The NetworkSecurityServer parameter on the IkeConfig statement specifies that the IKE daemon is configured to use network security services from an NSS server that is listening on IP address 9.1.1.1. Two NssStackConfig statements are shown. The ClientName parameters associate a local TCP/IP stack with an NSS client name. This is the name by which the NSS server knows this stack. The UserId parameter associates the client name with a user ID defined on the NSS server's system. Both the client name and user ID are used by the NSS server to verify that an NSS client is authorized to request certificate service, and to determine what certificates the client is authorized to use (For additional details, see “Steps for authorizing resources for NSS” on page 1086). The KeyRing parameter on the IkeConfig statement identifies the location of certificates for all stacks for which there are no NssStackConfig statements. The KeyRing parameter on the NssConfig statement identifies the location of certificates for all NSS client stacks that use the NSS server certificate service.

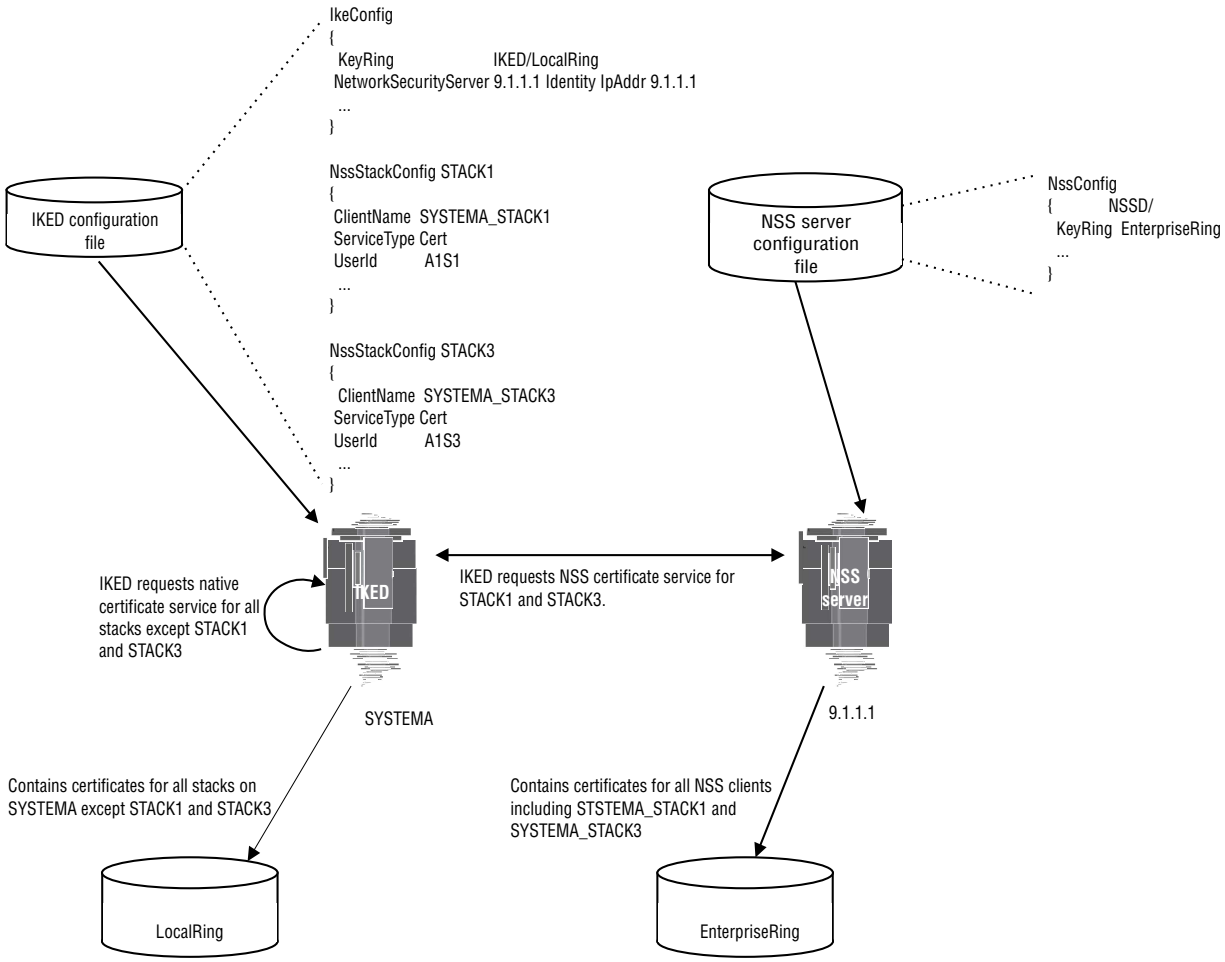


Figure 182. Partial configuration for the IKE daemon and an NSS server

The following subtopics provide steps for setting up the IKE daemon for RSA signature mode authentication:

- When the native certificate service is used
- When the certificate service of an NSS server is used

Tip: Use the following references to understand the concepts that are involved in using digital certificates with RACF:

- For more information about RACF key rings and digital certificates, see z/OS Security Server RACF Security Administrator's Guide.
- For more information about controlling the use of the RACDCERT command, see z/OS Security Server RACF Security Administrator's Guide.
- For a complete description of the facilities and authorizations that are needed to create and modify digital certificates and key rings, see z/OS Security Server RACF Command Language Reference.

Steps for setting up the IKE daemon for digital signature authentication when the native certificate service is used

You can configure the IKE daemon to use its own certificate service and to use RSA signature mode authentication. Details for each step are in the corresponding subtopics.

Procedure

Perform the following steps to set up the IKE daemon for digital signature authentication when the native certificate service is being used:

1. Define RACF facilities and access controls.
2. Define profiles to control access to the RACDCERT command.
3. Create a RACF key ring for the user ID under which the IKED is to run.
4. Install an X509 digital certificate to be used by the native IKE certificate service.

Step 1: Define RACF facilities and access controls

Procedure

To support RSA signature mode authentication in phase 1 negotiations, perform the following steps to give the IKE daemon the required access to a RACF key ring using the RDATALIB or FACILITY class. The RDATALIB class is preferred as it allows for granular authorization to a specified key ring. The FACILITY class allows global authorization to all key rings and certificates. See z/OS Security Server RACF Callable Services for additional information.

To provide access using the RDATALIB class:

1. If it is not already defined, create the RDATALIB definitions required to allow certificates to be stored and accessed from the RACF database by issuing the following TSO commands:

```
SETROPTS CLASSACT(RDATALIB) RACLIST(RDATALIB)
RDEFINE RDATALIB IKED.<keyRingName>.LST UACC(NONE)
```

2. To permit the IKED to the RDATALIB resource, issue the following TSO command:

```
PERMIT IKED.<keyRingName>.LST CLASS(RDATALIB) ID(IKED) ACCESS(READ)
```

3. Refresh the RDATALIB class:

```
SETROPTS RACLIST(RDATALIB) REFRESH
```

To provide access using the FACILITY class:

1. If they are not already defined, create the definitions that are required to allow certificates to be stored and accessed from the RACF database by issuing the following TSO commands:

```
RDEFINE FACILITY IRR.DIGTCERT.LISTRING UACC(NONE)
RDEFINE FACILITY IRR.DIGTCERT.LIST UACC(NONE)
```

2. To permit the IKED to the facilities, issue the following TSO commands:

```
PERMIT IRR.DIGTCERT.LIST CLASS(FACILITY) ID(IKED) ACC(READ)
PERMIT IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(IKED) ACC(READ)
```

3. Refresh the FACILITY class:

```
SETROPTS RACLIST(FACILITY) REFRESH
```

Step 2: Define profiles to control access to the RACDCERT command

Before you begin

The RACF database provides digital certificate and key ring support through the RACDCERT command. The administrator who is responsible for managing the RACF key ring for the IKED needs appropriate access to this command.

Procedure

Perform the following steps to define profiles to control access to this command:

1. If they are not already defined, create the definitions that are required to control access to the basic RACDCERT actions by issuing the following TSO commands:

```
RDEFINE FACILITY IRR.DIGTCERT.ADD UACC(NONE)
RDEFINE FACILITY IRR.DIGTCERT.ADDRING UACC(NONE)
RDEFINE FACILITY IRR.DIGTCERT.CONNECT UACC(NONE)
RDEFINE FACILITY IRR.DIGTCERT.GENCERT UACC(NONE)
RDEFINE FACILITY IRR.DIGTCERT.GENREQ UACC(NONE)
```

2. Issue the following TSO commands (where *userid* is the ID of the person who will be executing the RACDCERT command to manage digital certificates):

```
PERMIT IRR.DIGTCERT.ADD CLASS(FACILITY) ID(userid) ACC(CONTROL)
PERMIT IRR.DIGTCERT.ADDRING CLASS(FACILITY) ID(userid) ACC(UPDATE)
PERMIT IRR.DIGTCERT.CONNECT CLASS(FACILITY) ID(userid) ACC(CONTROL)
PERMIT IRR.DIGTCERT.GENCERT CLASS(FACILITY) ID(userid) ACC(CONTROL)
PERMIT IRR.DIGTCERT.GENREQ CLASS(FACILITY) ID(userid) ACC(CONTROL)
PERMIT IRR.DIGTCERT.LIST CLASS(FACILITY) ID(userid) ACC(CONTROL)
PERMIT IRR.DIGTCERT.LISTRING CLASS(FACILITY) ID(userid) ACC(UPDATE)
```

3. Refresh the FACILITY class:

```
SETROPTS RACLIST(FACILITY) REFRESH
```

4. If the RDATALIB class was used to control list access to the IKED key ring in Step 1 (Define RACF facilities and access controls), the administrator's ID should also be permitted to the RDATALIB profile.

```
PERMIT IKED.keyringname.LST CLASS(RDATALIB) ID(userid) ACC(UPDATE)
SETROPTS CLASSACT(RDATALIB) REFRESH
```

Step 3: Create a RACF key ring for the user ID under which the IKED is to run

Digital certificates are made available to the IKE daemon by connecting them to a key ring that is owned by the IKE daemon. To create a key ring for the IKE daemon, issue the following TSO command:

```
RACDCERT ID(IKED) ADDRING(ikeyring)
```

The value used for *ikeyring* is case sensitive.

Step 4: Install an X509 digital certificate to be used by the native certificate service

The procedure for installing a digital certificate on a SAF key ring for the native certificate service to use is similar to the procedure for installing a digital certificate on a SAF key ring for the NSS certificate service to use. For details, see “IPSec certificate management” on page 1373.

Steps for setting up the IKE daemon for digital signature authentication using the certificate service of an NSS server

You can configure the IKE daemon to use the certificate service of an NSS server and to use RSA signature mode authentication. Details for each step are in the corresponding subtopics.

Procedure

Perform the following steps to set up the IKE daemon for digital signature authentication when the certificate service of an NSS server is being used:

1. Update the IKE daemon configuration file to define NSS clients.
2. Install X509 digital certificates for NSS clients on the NSS server's key ring.
3. Authorize the NSS clients.
4. Enable HTTP Certificate Lookup (optional)

Step 1: Update the IKE daemon configuration file to define NSS clients

The IKE daemon configuration file must be updated to include the NetworkSecurityServer parameter and optionally the NetworkSecurityServerBackup parameter on the IkeConfig statement. These parameters identify the location of the NSS server. In addition, an NssStackConfig statement with a ServiceType value Cert must be added for each stack that will use the NSS certificate service. For additional details about configuring the IKE daemon, see z/OS Communications Server: IP Configuration Reference.

Step 2: Install X509 digital certificates for NSS clients on the NSS server's key ring

This step must be performed on the system where the NSS server will run.

The procedure for installing a digital certificates on a SAF key ring to be used by the NSS server is similar to the procedure used for installing a digital certificate on a SAF key ring for use by the native certificate service. For details, see “IPSec certificate management” on page 1373.

Step 3: Authorize the NSS clients

This step must be performed on the system where the NSS server will run.

NSS clients defined in the IKE daemon configuration file must be authorized to use the NSS server's certificate service. For authorization details, see step “8” on page 1088.

Step 4: Enable HTTP Certificate Lookup (optional)

The following actions occur during a phase 1 IKE exchange:

- Peers exchange encoded certificate information in certificate payloads
- The NSS server provides the IKED with certificate information to send
- The IKED forwards the encoded certificate information that it received from the NSS server

IKEv2 defines two new encoding types that require the use of an HTTP server:

- Hash and URL of a certificate
- Hash and URL of a certificate bundle

The NSS server supports these new encoding types; however, by default the IKED does not support sending or receiving them.

Use the CertificateURLLookupPreference parameter on the KeyExchangePolicy and KeyExchangeAction statements in the IP security policy configuration file to enable the IKED to use the hash and URL encoding types. Code the appropriate value on the CertificateURLLookupPreference parameter:

- Disallow, which is the default value, indicates that the IKED will not send the new encoding type and will ignore the new encoding type when received.
- Tolerate indicates that the IKED will not send the new encoding type, but it will accept the new encoding type when received.
- Allow indicates that the IKED may send the new encoding type and it will accept the new encoding type when received.

Rule: You must configure the NSS server appropriately before the IKED can send the new encoding types.

Enabling this capability might result in smaller messages being exchanged between the IKED and its remote security endpoint as well as the IKED and the NSS server; however, it may also result in increased latency during an IKEv2 negotiation.

For more details about the CertificateURLLookupPreference parameter on the KeyExchangePolicy and KeyExchangeAction statements, see z/OS Communications Server: IP Configuration Reference. For more details about configuring the NSS sever to use the new certificate encoding types, see “Using hash and URL certificate encoding types” on page 1099.

IPSec certificate management

The IKE daemon and NSS server require the ability to retrieve digital certificates from a RACF key ring, each associated with a particular identity, and also to perform operations with the associated private key. The IKED can own multiple certificates on its RACF key ring. The NSS server can own multiple certificates for multiple NSS clients (that is, stacks). You can install an X509 digital certificate in the following ways:

- Generate an X509 digital certificate and have it signed by a certificate authority.
- Generate a self-signed X509 digital certificate.
- Migrate an existing key database to a RACF key ring.

Steps for generating an X509 digital certificate and having it signed by a certificate authority

The IKE daemon and NSS server require the ability to retrieve digital certificates associated with a particular identity from a RACF key ring, and to perform operations with the associated private key.

Before you begin

Assume that you have an X509 digital certificate that has the X500 distinguished name `CN=SYSTEMA STACK1,OU=Inventory,O=IBM,C=US` and the domain name `ibm.com`. The certificate identifies the local IKE daemon that executes on z/OS with the user ID `IKED`.

Tip: If you are creating a certificate for a stack configured to use the certificate service from an NSS server, issue these commands against the RACF database for the system on which the NSS server runs. Modify the user ID in the examples to be the user ID that is running the NSS server and modify the key ring to be the key ring that is configured in the NSS server's configuration file.

Procedure

Perform the following steps to install the X509 digital certificate:

1. Generate a self-signed certificate for the server:

```
RACDCERT ID(IKED) GENCERT SUBJECTSDN(CN('SYSTEMA STACK1') OU('Inventory') O('IBM') C('US'))
                  WITHLABEL('SYSTEMA STACK1') ALTNAME(DOMAIN('ibm.com'))
```

2. Take one of the following actions:
   - RACF supplies certificates for many commercial certificate authorities. If you are using one of these supplied certificate authority certificates, follow the steps for supplied digital certificates in z/OS Security Server RACF Security Administrator's Guide.
   - If the certificate authority you are using is not one of the supplied certificate authorities, obtain the root certificate of the certificate authority that is to sign the certificate of the IKED, and place it in an MVS data set (for example, USER1.EXTCA1.CERT). Add it to the RACF database as follows:

     ```
     RACDCERT ID(IKED) ADD('USER1.EXTCA1.CERT') WITHLABEL('External CA') CERTAUTH
     ```
3. Create a certificate request to send to the chosen certificate authority.

   The certificate request that you create is based on the certificate that was created in step 1. Place this certificate into a data set called USER1.SYSTEMA.STACK1.GENREQ as follows:

   ```
   RACDCERT ID(IKED) GENREQ(LABEL('SYSTEMA STACK1'))
   DSN('USER1.SYSTEMA.STACK1.GENREQ')
   ```
4. Send the certificate request to the certificate authority.

   The certificate request is in base 64-encoded text. Typically, the request is sent to the certificate authority by cutting and pasting the certificate request into an email that is sent to the certificate authority.

   The certificate authority validates the certificate. If the certificate is approved by the certificate authority, it is signed by the certificate authority and returned to the requester.
5. Receive the returned certificate into a data set (for example, USER1.SYSTEMA.STACK1.CERT).

   The returned certificate is in base 64-encoded text. This can be done by cutting and pasting, with FTP, or with another technique.
6. Replace the self-signed certificate with the certificate that is signed by the certificate authority.

   The certificate is replaced only if the user ID that is specified as the ID value on the RACDCERT ADD command is the same user ID that was specified when the certificate was created. Ensure that the user ID is the same. Otherwise, the certificate is added, rather than replacing the self-signed certificate, and does not contain the certificate's private key.

   ```
   RACDCERT ID(IKED) ADD('USER1.SYSTEMA.STACK1.CERT') WITHLABEL('SYSTEMA STACK1')
   ```
7. Connect the certificate to an existing key ring:

   ```
   RACDCERT ID(IKED) CONNECT(LABEL('SYSTEMA STACK1') RING(ikeyring) USAGE(PERSONAL))
   ```
8. Connect the certificate authority's certificate to the key ring:

   ```
   RACDCERT ID(IKED) CONNECT(CERTAUTH LABEL('External CA') RING(ikeyring)
   USAGE(CERTAUTH))
   ```

   This completes the certificate hierarchy from root to SYSTEMA STACK1.
9. Add the following statement to the IKE daemon configuration file, iked.conf, or the NSS server configuration file, nssd.conf:

   ```
   Keyring   IKED/ikeyring
   ```

Results

You know you are done when the X509 digital certificate is available, and is mapped to the X500DN identity `CN=SYSTEMA STACK1,OU=Inventory,O=IBM,C=US` from the certificate's subject name, and the FQDN identity `ibm.com` from the certificate's alternate subject name.

You can verify that the certificates that you have created are connected to the key ring associated with user ID IKED by using the RACDCERT command and examining the output of the Ring Associations field. Verify that the certificate authority was created and added to the IKED/*ikeyring* as follows:

```
RACDCERT CERTAUTH LIST(LABEL('External CA')
```

Verify that the personal certificate for the IKE daemon was created and added to the IKED/*ikeyring* as follows:

```
RACDCERT ID(IKED) LIST(LABEL('SYSTEMA STACK1'))
```

Requirement: If the certificates connected to the key ring are for an NSS client, you must create a SERVAUTH profile for each certificate. You must give the user ID associated with the NSS client access to this profile. Create this profile in the RACF database for the system on which the NSS server runs. For details about these profiles, see "Steps for authorizing resources for NSS" on page 1086.

Steps for generating a self-signed X509 digital certificate

The IKE daemon and NSS server require the ability to retrieve digital certificates associated with a particular identity from a RACF key ring, and to perform operations with the associated private key.

Before you begin

The certificate that is assigned to the secure server is a locally-signed certificate rather than one signed by a certificate authority. Assume that the local certificate authority has the distinguished name of `OU='Local Certificate Authority',O=IBM,C=US`.

Requirement: If you are creating a certificate for a stack configured to use the certificate service from an NSS server, issue these commands against the RACF database for the system on which the NSS server runs. The user ID in the examples must be the user ID running the NSS server and the key ring must be the key ring configured in the NSS server's configuration file.

Procedure

Perform the following steps to implement a locally signed server certificate:

1. Generate a self-signed certificate to represent the local certificate authority:

   ```
   RACDCERT CERTAUTH GENCERT SUBJECTSDN(OU('Local Certificate Authority') O('IBM') C('US'))
   KEYUSAGE(CERTSIGN) WITHLABEL('IBM Local Certificate Authority')
   ```

   This certificate is used as the certificate authority certificate.
2. Export the certificate to a data set (in this case, USER1.LOCCERTA.CERT):

   ```
   RACDCERT CERTAUTH EXPORT(LABEL('IBM Local Certificate Authority')) DSN('USER1.LOCCERTA.CERT')
   ```
3. Create a certificate for the server that is signed with the certificate authority certificate that was created in step 1:

   ```
   RACDCERT ID(IKED) GENCERT SUBJECTSDN(CN('SYSTEMA STACK1') OU('Inventory') O('IBM') C('US'))
   WITHLABEL('SYSTEMA STACK1') ALTNAME(DOMAIN('ibm.com'))
   SIGNWITH(CERTAUTH LABEL('IBM Local Certificate Authority'))
   ```
4. Connect the certificate to an existing key ring:

   ```
   RACDCERT ID(IKED) CONNECT(LABEL('SYSTEMA STACK1') RING(ikeyring) USAGE(PERSONAL))
   ```
5. Connect the local certificate authority certificate to the key ring:

   ```
   RACDCERT ID(IKED) CONNECT(CERTAUTH LABEL('IBM Local Certificate Authority') RING(ikeyring)
   USAGE(CERTAUTH))
   ```

   This completes the certificate hierarchy from root to SYSTEMA STACK1.

6. Add the following statement to the IKE daemon configuration file, iked.conf, or the NSS server configuration file, nssd.conf:

```
Keyring   IKED/ikeyring
```

Results

You know you are done when the X509 digital certificate is available, and is mapped to the X500DN identity `CN=SYSTEMA STACK1,OU=Inventory,O=IBM,C=US` from the certificate's subject name, and the FQDN identity `ibm.com` from the certificate's alternate subject name.

You can verify that the certificates that you have created are connected to the key ring associated with user ID IKED by using the RACDCERT command and examining the output of the `Ring Associations` field. Verify that the certificate authority was created and added to the IKED/*ikeyring* as follows:

```
RACDCERT CERTAUTH LIST(LABEL('IBM Local Certificate Authority'))
```

Verify that the personal certificate for the IKE daemon was created and added to the IKED/*ikeyring* as follows:

```
RACDCERT ID(IKED) LIST(LABEL('SYSTEMA STACK1'))
```

Requirement: If the certificates connected to the key ring are for an NSS client, you must create a SERVAUTH profile for each certificate. You must give the user ID associated with the NSS client access to this profile. Create this profile in the RACF database for the system on which the NSS server runs. For details about these profiles, see “Steps for authorizing resources for NSS” on page 1086.

Steps for migrating an existing key database to a RACF key ring

The IKE daemon and NSS server require the ability to retrieve digital certificates associated with a particular identity from a RACF key ring, and to perform operations with the associated private key.

Before you begin

To migrate an existing key database file to a RACF key ring, see the information on migrating key database files to RACF key rings in z/OS Cryptographic Services System SSL Programming.

Procedure

Perform the following steps to migrate keys and certificates that are stored in an existing z/OS key database into a RACF key ring:

1. Using gskkyman, export the certificate and private key to a password-protected PKCS#12 file. For details on copying a certificate with its private key, see z/OS Cryptographic Services System SSL Programming.
2. Copy the newly created PKCS#12 file to an MVS data set.
3. Use the RACDCERT command with the ADD operand to define a certificate and private key. The data set name that was created in step 2 contains the certificate.
4. Use the RACDCERT command with the ADDRING operand to create a new key ring in RACF.
5. Use the RACDCERT command with the CONNECT operand to add the certificate and private key to one or more existing RACF key rings.

Appendix F. Using an LDAP server for policy definitions

Lightweight Directory Access Protocol (LDAP) is a fast-growing technology for accessing common directory information. LDAP has been embraced and implemented in most network-oriented middleware. As an open, vendor-neutral standard, LDAP provides an extendable architecture for centralized storage and management of information that needs to be available for today's distributed systems and services.

Note: If the z/OS LDAP server is used, a Db2 backend is required.

Policy object model overview

Policies consist of several related objects. The main object is the *policy rule*. A policy rule object refers to one or more *policy condition*, *policy action*, or *policy time period condition* objects, and also contains information on how these objects are to be used. Policy time period objects are used to determine when a given policy rule is active. Active policy objects are related in a way that is analogous to an 'IF' statement in a program. For example:

```
IF condition THEN action
```

In other words, when the set of conditions referred to by a policy rule are TRUE, then the policy actions associated with the policy rule are executed.

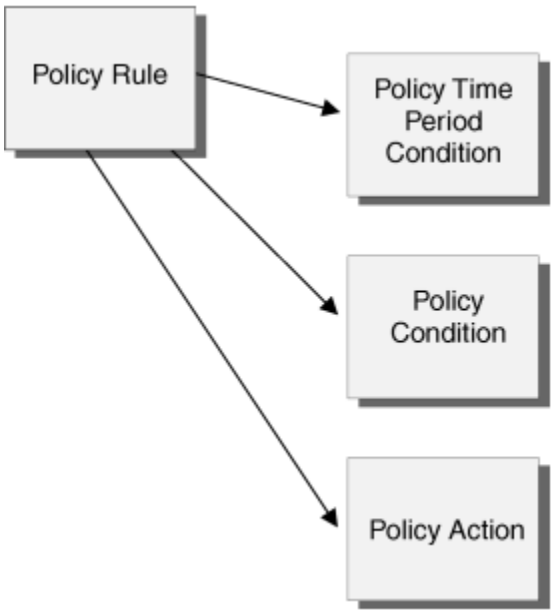


Figure 183. Basic policy objects

Policy rules can refer to one or more policy conditions. A policy rule with a single policy condition is known as a *simple* rule, and one with more conditions is known as a *complex* rule. Complex policy rules can have their conditions evaluated according to one of two different methods. The first is Disjunctive Normal Form (DNF), which means an ORed set of ANDed conditions. The second is Conjunctive Normal Form (CNF), which means an ANDed set of ORed conditions. In order to accomplish these evaluations, individual policy conditions are assigned an arbitrary group number, and also an indication of whether or not the condition is negated. For example, consider the following set of conditions for a policy rule:

```
C1:  Group Number = 1, Condition Negated = FALSE
C2:  Group Number = 1, Condition Negated = TRUE
C3:  Group Number = 1, Condition Negated = FALSE
```

© Copyright IBM Corp. 2000, 2022

```
C4:  Group Number = 2, Condition Negated = FALSE
C5:  Group Number = 2, Condition Negated = FALSE
```

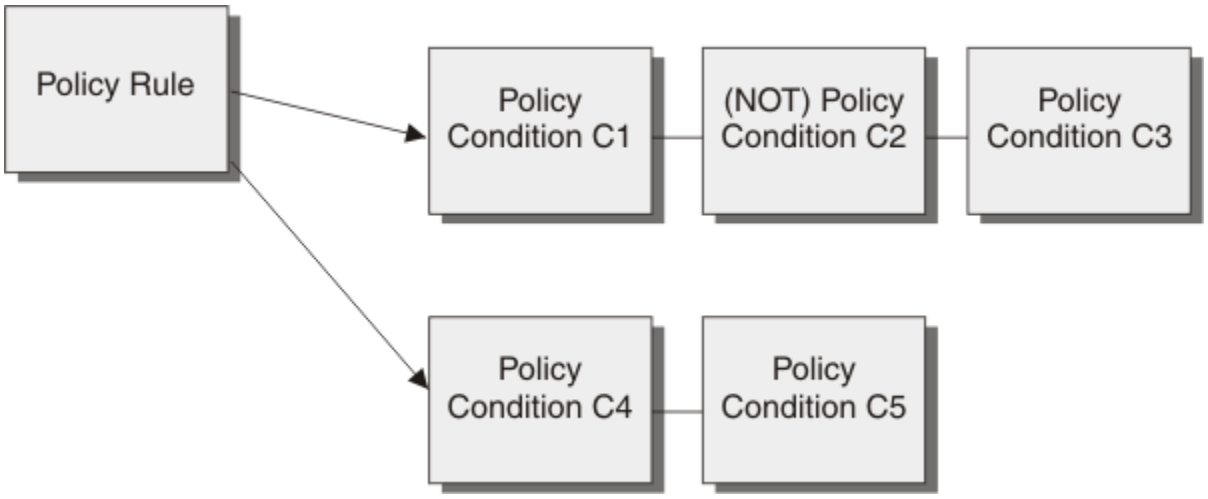


Figure 184. Complex policy conditions

If the conditions are to be evaluated using DNF, then the overall condition for the policy rule is:

```
(C1 AND (NOT C2) AND C3) OR (C4 AND C5)
```

On the other hand, if CNF is used to evaluate the conditions, then the overall condition for the policy rule is:

```
(C1 OR (NOT C2) OR C3) AND (C4 OR C5)
```

Complex rules can be split into multiple simple rules. Negated conditions are not allowed in a rule if explosion is to be performed. Consider the following set of conditions for a policy rule:

```
C1:  Group Number = 1, Condition Negated = FALSE
C2:  Group Number = 1, Condition Negated = FALSE
C3:  Group Number = 1, Condition Negated = FALSE
C4:  Group Number = 2, Condition Negated = FALSE
C5:  Group Number = 2, Condition Negated = FALSE
```

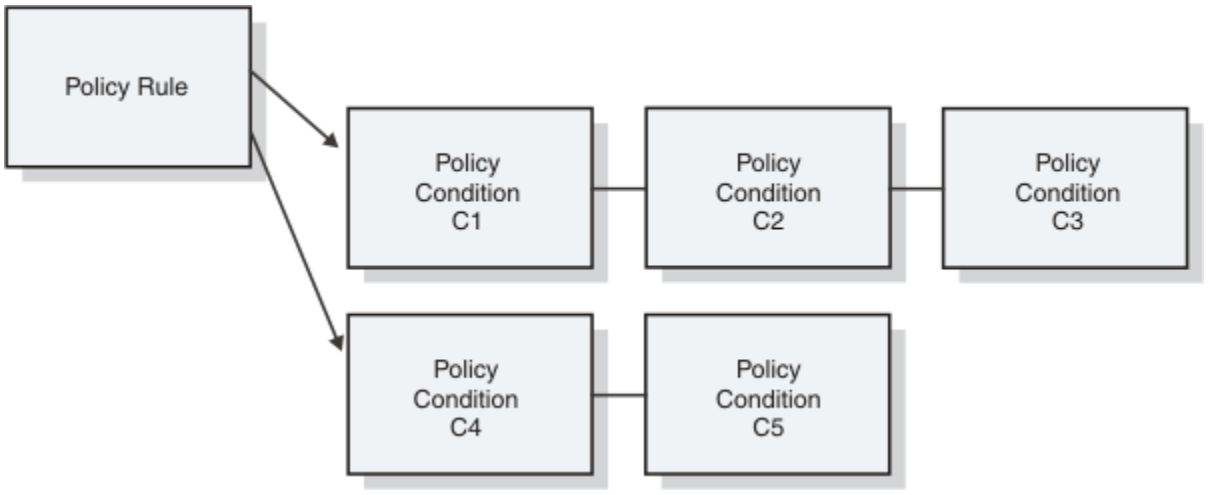


Figure 185. Complex policy conditions before explosion

If DNF is used to evaluate the conditions, splitting the complex rule produces the following simple rules:

```
Simple Rule 1:  C1 AND C2 AND C3
Simple Rule 2:  C4 AND C5
```

If CNF is used to evaluate the conditions, splitting the complex rule produces the following simple rules:

```
Simple Rule 1:  C1 AND C4
Simple Rule 2:  C1 AND C5
Simple Rule 3:  C2 AND C4
Simple Rule 4:  C2 AND C5
Simple Rule 5:  C3 AND C4
Simple Rule 6:  C3 AND C5
```

Policy actions specify actions to take when the set of conditions for a policy rule evaluate to TRUE. The policy model allows multiple actions for a policy rule. Many policy rules typically use only a single action, but multiple actions make sense for some policy types.

Policy conditions and actions can either be specific to a single rule, or be reusable among several policy rules. To allow either type of conditions and actions, and to specify related information such as condition group number and negation indicator, several other policy objects are required. First are *policy condition association* and *policy action association* objects. These objects contain condition and action related attributes, respectively, and may directly contain policy conditions and actions (rule-specific).

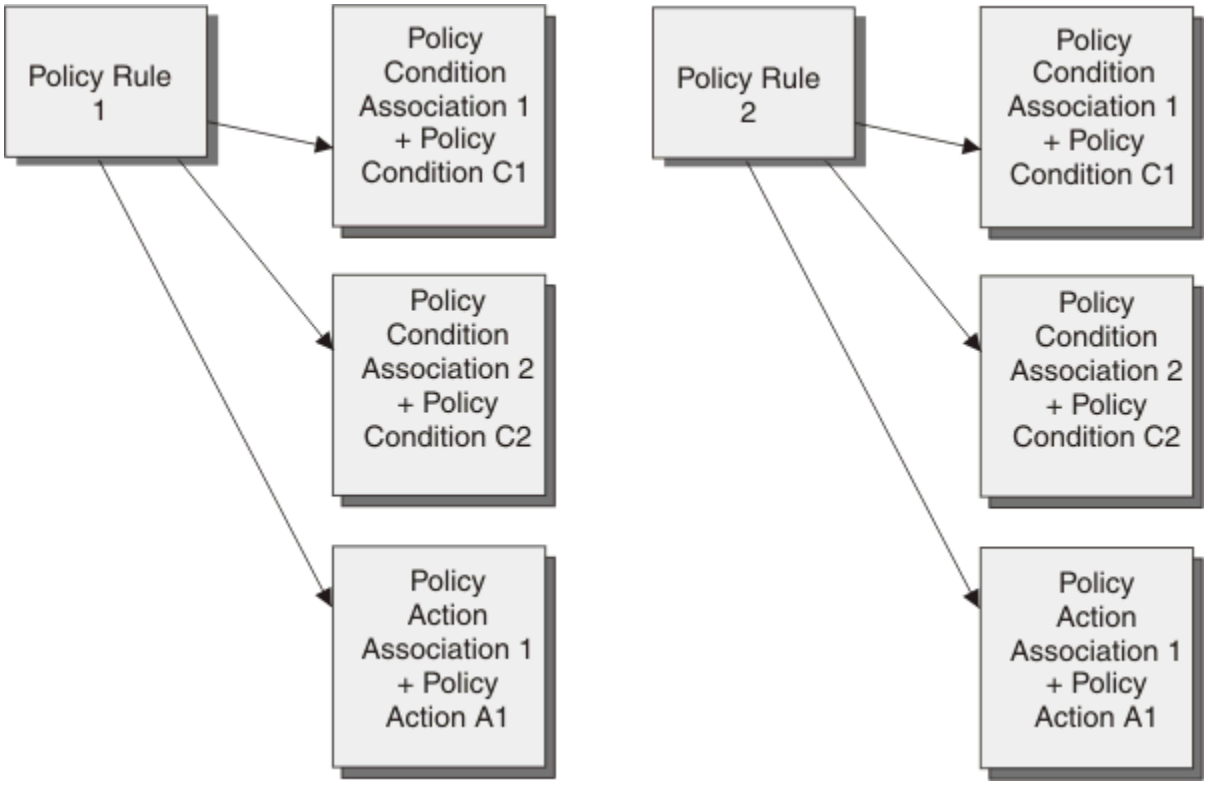


Figure 186. Rule-specific conditions and actions

The policy association objects alternatively may refer to conditions and actions (reusable). *Policy condition instance* and *policy action instance* objects are used to represent reusable policy conditions and actions, respectively.

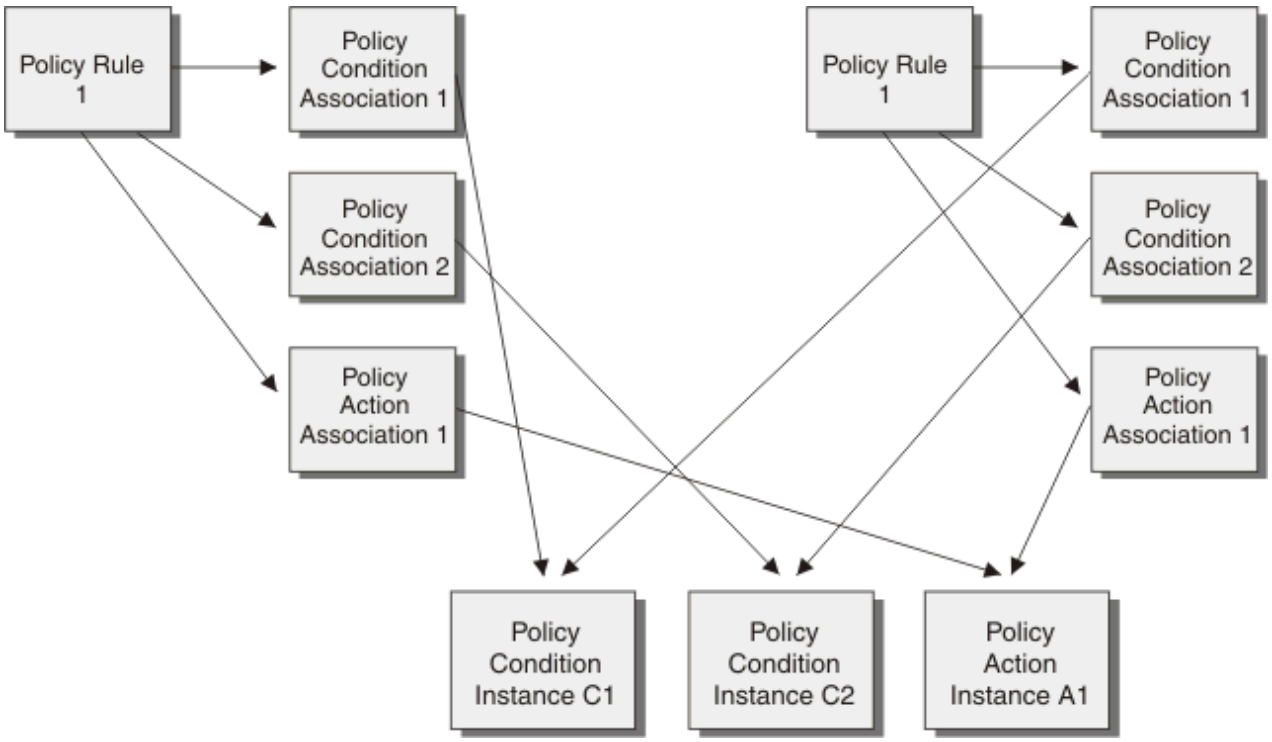


Figure 187. Reusable conditions and actions

Primarily for administrative grouping of policy rules, the *policy group* object is used. Policy groups can refer either to policy rules or to policy groups. This allows related policy rules to be grouped together, and also allows policy groups to be grouped to any needed level of nesting.

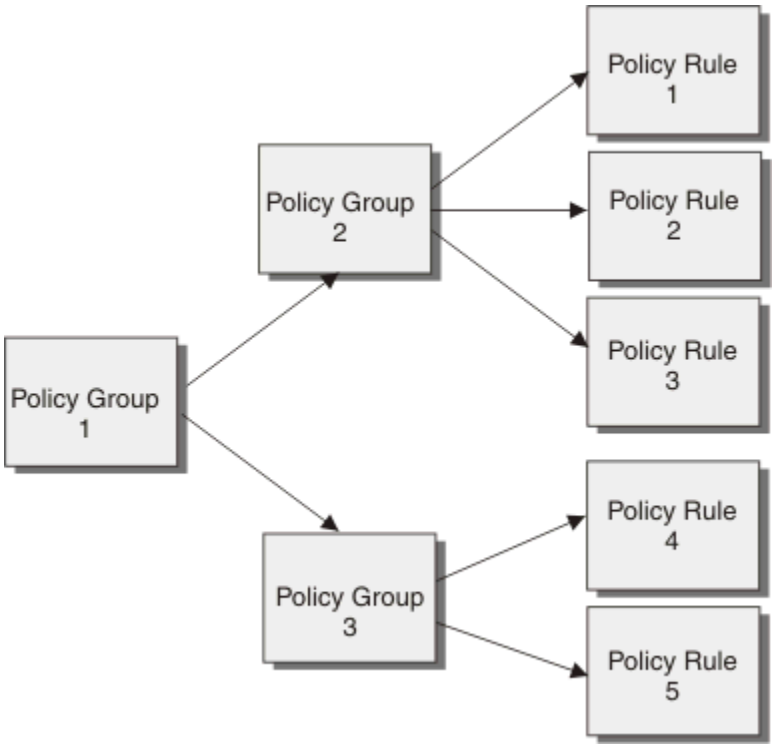


Figure 188. Policy groups

Overview of the object classes

Policies defined on an LDAP server are comprised of one or more objects, each with a defined object class, and a unique name. Object names are in the form of LDAP Distinguished Names (DNs), which are text strings composed of individual parts known as Relative Distinguished Names (RDNs). The structure of the naming defines a hierarchical tree, in a manner similar to directories in a hierarchical file system. For example, consider the following set of objects:

```
Object 1, DN: o=IBM, c=US
Object 2, DN: cn=group_1, o=IBM, c=US
Object 3, DN: cn=group_5, o=IBM, c=US
Object 4, DN: cn=group_1_sub_A, cn=group_1, o=IBM, c=US
```

This set of objects can be viewed as a tree, with Object 1 as the root. Objects 2 and 3 are branches under the root, with Object 4 a branch under Object 2. The names used are attributes of the objects they define. For example, Object 2, whose name starts with "cn=group_1" contains a cn attribute with the value group_1. See z/OS IBM Tivoli Directory Server Administration and Use for z/OS for more information on LDAP naming.

Object class names define the type of each LDAP object. The *top* object class is predefined and is the root of all other object classes.

There are three types of object classes.

Abstract object classes
: Used to define broad concepts, such as policy and to define basic attributes that apply to all subclasses.

Structural object classes
: Basic building blocks, and are the only type of object class that can be instantiated as real objects on an LDAP server.

Auxiliary object classes
: Used to define attributes that are shared among different structural object classes, or are used to extend the basic set of objects.

Attributes from auxiliary classes are *attached* to structural objects by including them in the structural objects, and also by including the auxiliary object class as one of the values of the objectClass attribute in the structural object.

The following object classes are recognized by the Policy Agent. The indentation defines subclasses. For example, ibm-policyGroup is a subclass of ibm-policy, and therefore inherits all of the attributes defined for ibm-policy.

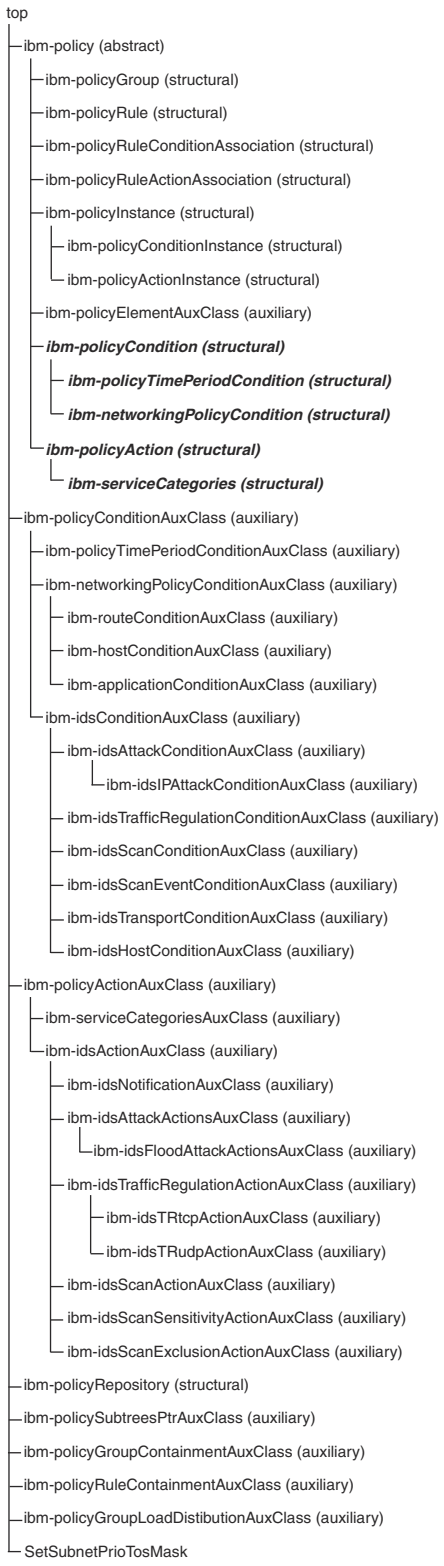


Figure 189. LDAP schema object class hierarchy

Note: The classes identified in bold italics in Figure 189 on page 1382 are for schema version 2 and are mutually exclusive with the schema version 3 classes with similar names.

Object class name	Purpose of object
Top	Used to anchor the LDAP hierarchical tree root.

Object class name	Purpose of object
ibm-policy	Used as the root for all policy objects.
ibm-policyGroup	Defines a policy group object.
ibm-policyRule	Defines a policy rule object.
ibm-policyRuleConditionAssociation	Defines an association between a policy rule object and a policy condition.
ibm-policyRuleActionAssociation	Defines an association between a policy rule object and a policy action.
ibm-PolicyInstance	Defines an instance of a reusable policy object.
ibm-PolicyConditionInstance	Defines an instance of a reusable policy condition object.
ibm-PolicyActionInstance	Defines an instance of a reusable policy action object.
ibm-PolicyElementAuxClass	Defines an auxiliary class that can be used to *tag* non-policy objects as though they were policy objects.
ibm-policyCondition	Defines a policy condition object. (schema version 2 — supported for migration)
ibm-policyTimePeriodCondition	Defines an auxiliary class to represent time periods during which a policy rule is considered to be active. (schema version 2 — supported for migration)
ibm-networkingpolicyCondition	Defines a subclass of ibm-PolicyCondition used to define networking policy conditions. (schema version 2 — supported for migration)
ibm-policyAction	Defines a policy action object. (schema version 2 — supported for migration)
ibm-serviceCategories	Defines an auxiliary class to represent a set of QoS attributes for a policy action. (schema version 2 — supported for migration)
ibm-policyConditionAuxClass	Defines an auxiliary class for generic policy conditions.
ibm-policyTimePeriodConditionAuxClass	Defines an auxiliary class to represent time periods during which a policy rule is considered to be active.
ibm-networkingPolicyConditionAuxClass	Defines an auxiliary class used to define networking policy conditions.
ibm-routeConditionAuxClass	Defines an auxiliary class to represent QoS routing conditions for a policy rule.
ibm-hostConditionAuxClass	Defines an auxiliary class to represent QoS host (end-point) conditions for a policy rule.
ibm-applicationConditionAuxClass	Defines an auxiliary class to represent QoS application and transport conditions for a policy rule.

Object class name	Purpose of object
ibm-idsConditionAuxClass	Defines an auxiliary class to represent generic IDS conditions.
ibm-idsAttackConditionAuxClass	Defines an auxiliary class to represent IDS attack conditions.
ibm-idsIPAttackConditionAuxClass	Defines an auxiliary class to represent IDS IP attack conditions.
ibm-idsTrafficRegulationConditionAuxClass	Defines an auxiliary class to represent IDS Traffic Regulation conditions.
ibm-idsScanConditionAuxClass	Defines an auxiliary class to represent IDS global scan conditions.
ibm-idsScanEventConditionAuxClass	Defines an auxiliary class to represent IDS scan event conditions.
ibm-idsTransportConditionAuxClass	Defines an auxiliary class to represent IDS transport conditions.
ibm-idsHostConditionAuxClass	Defines an auxiliary class to represent IDS host conditions.
ibm-policyActionAuxClass	Defines an auxiliary class for generic policy actions.
ibm-serviceCategoriesAuxClass	Defines an auxiliary class to represent a set of QoS attributes for a policy action.
ibm-idsActionAuxClass	Defines an auxiliary class to represent generic IDS actions.
ibm-idsNotificationAuxClass	Defines an auxiliary class to represent notification options for IDS actions.
ibm-idsAttackActionsAuxClass	Defines an auxiliary class to represent attack attributes for IDS actions.
ibm-idsFloodAttackActionsAuxClass	Defines an auxiliary class to represent flood-specific attack attributes for IDS actions.
ibm-idsTrafficRegulationActionAuxClass	Defines an auxiliary class to represent generic Traffic Regulation attributes for IDS actions.
ibm-idsTRtcpActionAuxClass	Defines an auxiliary class to represent Traffic Regulation TCP attributes for IDS actions.
ibm-idsTRudpActionAuxClass	Defines an auxiliary class to represent Traffic Regulation UDP attributes for IDS actions.
ibm-idsScanActionAuxClass	Defines an auxiliary class to represent global scan attributes for IDS actions.
ibm-idsScanSensitivityActionAuxClass	Defines an auxiliary class to represent scan sensitivity attributes for IDS actions.
ibm-idsScanExclusionActionAuxClass	Defines an auxiliary class to define scan exclusion lists for IDS actions.
ibm-policyRepository	Defines a repository for generic reusable policy objects.

Object class name	Purpose of object
ibm-policySubtreesPtrAuxClass	Defines an auxiliary class to represent pointers to subtrees in the LDAP directory tree to be retrieved by the Policy Agent. This allows entire subtrees to be retrieved at once, improving retrieval performance in some situations.
ibm-policyGroupContainmentAuxClass	Defines an auxiliary class for binding a policy group object to another policy group.
ibm-policyRuleContainmentAuxClass	Defines an auxiliary class for binding a policy rule object to another policy group.
ibm-policyGroupLoadDistributionAuxClass	Defines an auxiliary class to represent load distribution attributes for policy groups. The load distribution attributes are applied to all policy rules that are pointed to by groups to which this auxiliary class has been attached.
SetSubnetPrioTosMask	Defines a mapping of outbound IPv4 packet Type of Service (ToS) byte or IPv6 packet Traffic Class values to QDIO device priorities and Virtual LAN (VLAN) user priorities.

Policy objects are used to accomplish the following objectives:

- Group related objects together. Policy groups can contain related policy rules, and can also contain other policy groups. This allows objects to be grouped in various administrative ways. If the resulting objects will be retrieved by any Policy Agent prior to z/OS Communications Server V1R2, then the object should not include the values ibm-policyGroupContainmentAuxClass, ibm-policyRuleContainmentAuxClass or ibm-policyGroupLoadDistributionAuxClass for the objectClass attribute.
- Specify conditions for a policy rule. The conditions are used to filter traffic packets, and can specify attributes such as source and destination port, application name, protocol, and so on. Policy rules can be either simple or complex, depending on the nature of the specified conditions. When a single condition is specified, the rule is a simple rule. This single condition can be composed of any of the condition attributes, but only one instance of each. For example, only a single destination port range can be specified in a simple rule. Complex rules specify more than one condition. The specified conditions are organized into one or more levels, and each level is composed of one or more conditions. Each condition can be composed of one instance of any of the condition attributes. The conditions can thus be thought of as a two-dimensional array. Any individual condition can be negated. Two types of processing are applied to the conditions, depending on the specified condition list type:
  - Disjunctive Normal Form (DNF). DNF conditions are logically ANDed at each level, and ORed between levels.
  - Conjunctive Normal Form (CNF). CNF conditions are logically ORed at each level, and ANDed between levels.

  For example, suppose five conditions are specified, two at one level and three at another:

```
Level 1: C1, NOT C2
Level 2: C3, C4, C5
```

  If DNF is specified, the conditions are evaluated as:

```
(C1 AND NOT C2) OR (C3 AND C4 AND C5)
```

  CNF evaluation of the same conditions is:

```
(C1 OR NOT C2) AND (C3 OR C4 OR C5)
```

This allows a wide variety of conditional logic to be defined for policy rules.

- Specify time periods during which policy rules are active. Active policy rules are those that are installed in a TCP/IP stack by the Policy Agent. A wide variety of attributes are available to specify time periods, and up to 25 time periods can be specified for any policy rule. The policy rule is active if any of the specified time intervals include the current time.
- Specify actions to take on behalf of traffic that maps to active policy rules, based on the evaluation of its conditions. QoS Actions contain a scope attribute that indicates the type of policy being defined, namely Differentiated Services, RSVP, or both Differentiated Services and RSVP. Up to four actions can be specified for each rule, but only one action per scope can be active at a time. IDS actions contain an action type that indicates the type of policy being defined, namely Attack, TR, Scan Global, or Scan Event. Only one IDS action can be specified for each rule. QoS and IDS actions (or conditions) can't be mixed within a single policy rule.

Considerations for defining LDAP objects

LDAP objects can refer to other objects, using the DN of the referenced object. For example, a policy rule can be separated from its conditions and time periods, with those objects being referenced by the rule object.

Each LDAP object is composed of a number of attributes. Some of the attributes are generic LDAP attributes that apply to all LDAP objects. Other attributes are used only for Version 1 policy definitions. All other Version 2 and later policy attributes must begin with a unique prefix:

```
ibm-
```

When defining complex policy rules (those with more than one condition or action), two mutually exclusive methods can be used to associate the conditions or actions with the rule:

- The ibm-policyConditionListDN and ibm-policyActionListDN attributes can be omitted from the rule. In this case, the condition and action association objects MUST be created as subordinate objects to the policy rule, in other words, under the rule in the directory subtree. This is known as Directory Information Tree (DIT)-containment.
- The ibm-policyConditionListDN and ibm-policyActionListDN attributes can be specified in the rule. In this case, the condition and action association objects SHOULD be created as subordinate objects to the policy rule, in other words, under the rule in the directory subtree. However, this is not a requirement, only a recommendation. The objects can actually exist anywhere in the DIT.

Policy Agent retrieval of LDAP objects

The design of the LDAP object tree should be carefully thought out. The Policy Agent uses a variety of mechanisms to search for and retrieve objects from an LDAP server:

- An initial search is done for a subtree of objects based on the SearchBaseDN parameter on the ReadFromDirectory statement.
- If any objects retrieved by this initial search contain subtree pointer references (using the ibm-policySubtreesAuxContainedSet attribute) then a search is done for all such subtrees. This is a recursive search: additional objects retrieved might also contain subtree pointer references.
- The above searches use a filter to retrieve only certain object classes. For LDAP protocol version 3, the default is to scan only for the ibm-policy object class. This is an abstract object class from which all other policy object classes are derived. Most LDAPv3 servers implement abstract and auxiliary classes such that this search will properly retrieve policy, and only policy, object classes. However, some LDAPv3 servers do not honor abstract/auxiliary object classes as a search filter. For these servers, specify LDAP_AbstractPolicy NO on the ReadFromDirectory statement. This causes the searches to use a filter that retrieves ALL object classes.

- All of the above searches may be scoped, or filtered, using keywords specified on the ReadFromDirectory statement parameters SearchPolicyKeyword, SearchPolicyGroupKeyword, or SearchPolicyRuleKeyword. The LDAP server returns only objects with any matching keywords.
- Some objects retrieved using the above searches may contain DN pointer references to additional objects. These objects are individually retrieved. If the object to be retrieved is a policy rule, then a subtree search is performed, using the keywords specified on the ReadFromDirectory statement. All other objects are retrieved as single objects, using the DN pointers (no keywords are used on the search).
- All policy rule objects retrieved using the above searches are further filtered using the PolicyRole parameter on the ReadFromDirectory statement. Any rules that do not match policy roles specified on the ReadFromDirectory statement are discarded.

Therefore, it is possible to design an LDAP tree such that a minimal set of objects is initially retrieved, followed by many additional individual LDAP retrievals. If the total set of objects is large, there is a performance impact to retrieving objects in this manner. If possible, try to design the tree and the ReadFromDirectory parameters to retrieve the largest set of objects initially, to achieve the best performance, or to use subtree pointer references to retrieve larger sets of objects in one operation.

LDAP sample files

The following set of sample files provides an example of policy definitions in LDAP. For more information on using these sample files, see "Using the sample LDAP objects" on page 1388.

/usr/lpp/tcpip/samples/pagent.ldif
This file contains the top level directory structure for the set of sample quality of service (QoS) and intrusion detection services (IDS) policies.

/usr/lpp/tcpip/samples/pagent_starter_QOS.ldif
This file contains the starter set sample of LDAP definitions of QoS objects. This file requires the directory structure defined in sample file pagent.ldif.

/usr/lpp/tcpip/samples/pagent_starter_IDS.ldif
This file contains the starter set sample of LDAP definitions of IDS objects. This file requires the directory structure defined in sample file pagent.ldif.

/usr/lpp/tcpip/samples/pagent_advanced_QOS.ldif
This file contains the advanced set sample of LDAP definitions of QoS objects. This file requires the objects defined in the QoS starter set sample file pagent_starter_QOS.ldif and the directory structure defined in sample file pagent.ldif.

/usr/lpp/tcpip/samples/pagent_advanced_IDS.ldif
This file contains the advanced set sample of LDAP definitions of IDS objects. This file requires the objects defined in the IDS starter set sample file pagent_starter_IDS.ldif and the directory structure defined in sample file pagent.ldif.

The next set of samples is the definition of the schemas in LDAP protocol version 3 format. They must be installed on the LDAPv3 server in the correct order as an object in the server's database, rather than as configuration information. This process is known as schema publication. See RFCs 1804 and 2251. The files need to be specified on ldapmodify commands to modify the cn:schema entry in the server's database. See "Installing the schema definition on the LDAP server" on page 1388 for more information.

/usr/lpp/tcpip/samples/pagent_r8qosschema.ldif
This file contains the schema version 3 QoS schema definitions.

/usr/lpp/tcpip/samples/pagent_r5idsschema.ldif
This file contains the schema version 3 IDS schema definitions.

The next set of samples contain the text of draft documents used to develop the Version 3 schema.

/usr/lpp/tcpip/samples/pagent_pcim.txt
This file contains the draft version of the proposed Policy Core Information Model (PCIM) used as the basis for the support in this release. PCIM is described by RFC 3060 but this file is an earlier draft level.

/usr/lpp/tcpip/samples/pagent_core.txt
This file contains the draft version of the proposed Policy Core LDAP Schema Internet Draft used in z/OS V1R2 and later releases.

Note: This file is provided as-is and there are some differences between the draft and the implementation. The intent of this file is to provide background information on the level of the supported schema.

/usr/lpp/tcpip/samples/pagent_cond.txt
This file contains the draft version of the proposed Policy Conditions Internet Draft used in z/OS V1R2 and later releases.

Note: This file is provided as-is and there are some differences between the draft and the implementation. The intent of this file is to provide background information on the level of the supported schema.

Installing the schema definition on the LDAP server

The files that define the schema supported by the Policy Agent are shipped as a set of sample files. You need to modify the configuration of the LDAP server to include these schema definition files.

For LDAP protocol version 3, the schema definition is shipped in *ldif* format and installed on the LDAP server as a modification to the generic schema entry, known as a subschema. You must modify the existing schema entry to include the supported schema as a subschema by using the **ldapmodify** command. The schema definition files that you must install are located in the /usr/lpp/tcpip/samples directory. You must install the files in the following order:

1. pagent_r8qosschema.ldif
2. pagent_r5idsschema.ldif

This process is supported for the z/OS LDAP server.

To install the schema definitions, use commands like those shown in the following examples:

```
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_r8qosschema.ldif
```

```
ldapmodify -h <server address> -p <server port> -D <administrator userid>
-w <password> -f /usr/lpp/tcpip/samples/pagent_r5idsschema.ldif
```

See the TDBM backend information in z/OS IBM Tivoli Directory Server Administration and Use for z/OS for more details.

Using the sample LDAP objects

There are 5 sample files that provide examples of policy definitions in LDAP:

- pagent.ldif
- pagent_starter_IDS.ldif
- pagent_starter_QOS.ldif
- pagent_advanced_IDS.ldif
- pagent_advanced_QOS.ldif

For brief descriptions of these files, see “Policy sample files” on page 800. You can either use some or all of these predefined policies in the starter and advanced sets, or modify them as needed.

The recommended way to create customized policies is to copy the sample policies you want to change to the custom portion of the pagent.ldif file (under the appropriate cn=custom root, QoS or IDS), modify them as needed, and then point to the custom set as the search base on the ReadFromDirectory statement.

For example, the pagent.ldif file has the following hierarchical structure [this shows the relevant parts of the Distingushed Name (DN) for each object]:

```
o=IBM, c=US (root object)
    cn=repository (root of all reusable policy conditions and actions)
    ou=policy (root of all policy groups and rules)
       cn=groups (root of sample groups)
       cn=starter (root of simple starter set of policies)
           cn=IDS (IDS starter set - actually defined in file
pagent_starter_IDS.ldif)
           cn=QoS (QOS starter set - actually defined in file
pagent_starter_QOS.ldif)
       cn=advanced (root of advanced set of policies)
           cn=IDS (IDS advanced set - actually defined in file
pagent_advanced_IDS.ldif)
           cn=QoS (QOS advanced set - actually defined in file
pagent_advanced_QOS.ldif)
       cn=custom (root of customer-defined set of policies)
           cn=IDS (root of customer-defined IDS policies (empty))
           cn=QoS (root of customer-defined QOS policies (empty))
```

To obtain only the customized policies, specify the top custom policy group object as the search base on the ReadFromDirectory statement as follows:

```
ReadFromDirectory {
...
SearchPolicyBaseDN   dn:cn=custom, ou=policy, o=IBM, c=US
...
}
```

Note: If your LDAP server has a root structure other than "o=IBM, c=US", be sure to change the root structure in all the files you want to use by replacing every instance of "o=IBM, c=US" with the appropriate root used on your LDAP server.

Defining QoS policies using LDAP

This topic contains examples for defining QoS policies using LDAP.

Differentiated Services policy example

The goal of this Differentiated Services policy is to map a subset of the traffic outbound from an FTP server.

This policy is identified as a Differentiated Services policy by the ibm-PolicyScope:DataTraffic attribute in the ibm-PolicyActionInstance object.

The following statements apply to the example in this topic:

- The policy rule selects traffic originated by ports in the range 20-21 (FTP outbound data connection uses port 20) from the source address 200.50.23.11 or 200.50.33.14 or 202::B055:1.
- The policy rule is active on weekdays between 6 a.m. and 10 p.m. local time, between the dates 7/1/2000 and 7/1/2005.
- The policy action specifies that the ToS byte be set to '10000000' for traffic that conforms to this policy.
- The action establishes a *token bucket* traffic conditioner with a mean rate of 256 kilobits per second, a peak rate of 512 kilobits per second, and a burst size of 64 kilobytes. Any traffic that exceeds these specifications will be sent as *best effort*, with an accompanying ToS byte of '00000000'.

```
dn:cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:diffserv-rule
ibm-policyRuleName:diffserv-rule
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListType:2
ibm-policyRuleConditionListDN:cn=condassoc1, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc2, cn=diffserv-rule, cn=QoS, cn=advanced,
```

```
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc3a, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc3b, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleConditionListDN:cn=condassoc3c, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleActionListDN:cn=actassoc1, cn=diffserv-rule, cn=QoS, cn=advanced,
    ou=policy, o=IBM, c=US
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:10
ibm-policyRuleMandatory:TRUE
ibm-policyRuleSequencedActions:1
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Differentiated Services rule

dn:cn=condassoc1, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:diffserv-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable CNF condition at level 1 - TCP

dn:cn=condassoc2, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:diffserv-condition2
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=ftpdPorts, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable CNF condition at level 2 - ftpd ports

dn:cn=condassoc3a, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3a
ibm-policyConditionName:diffserv-condition3a
ibm-policyConditionGroupNumber:3
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=Host1, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents first reusable CNF condition at level 3 - host1

dn:cn=condassoc3b, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3b
ibm-policyConditionName:diffserv-condition3b
ibm-policyConditionGroupNumber:3
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=Host2, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents second reusable CNF condition at level 3 - host2

dn:cn=condassoc3c, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3c
ibm-policyConditionName:diffserv-condition3c
ibm-policyConditionGroupNumber:3
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=Host3, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents third reusable CNF condition at level 3 - host3

dn:cn=actassoc1, cn=diffserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:diffserv-action
```

```
ibm-policyActionOrder:1
ibm-policyActionDN:cn=tokenbucket, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - token bucket

dn: cn=tokenbucket, cn=QoSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:tokenbucket
ibm-policyActionName:tokenbucket-action
ibm-PolicyScope:DataTraffic
ibm-OutgoingTOS:10000000
ibm-DiffServInProfileRate:256
ibm-DiffServInProfilePeakRate:512
ibm-DiffServInProfileTokenBucket:512
ibm-DiffServInProfileMaxPacketSize:120
ibm-DiffServOutProfileTransmittedTOSByte:00000000
ibm-DiffServExcessTrafficTreatment:BestEffort
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS diffserv token bucket action

dn:cn=ftpdPorts, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:ftpdPorts
ibm-policyConditionName:ftpdPorts-condition
ibm-sourcePortRange:20-21
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS ftpd ports condition

dn:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP-condition
ibm-protocolNumberRange:6
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS protocol TCP condition

dn:cn=Host1, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
cn:Host1
ibm-policyConditionName:Host1-condition
ibm-SourceIPAddressRange:3-200.50.23.11
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS Host 1 condition

dn:cn=Host2, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
cn:Host2
ibm-policyConditionName:Host2-condition
ibm-SourceIPAddressRange:3-200.50.33.14
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS Host 2 condition

dn:cn=Host3, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
cn:Host3
ibm-policyConditionName:Host3-condition
ibm-SourceIPAddressRange:5-202::B055:1
```

```
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS Host 3 condition

dn:cn=period1, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period1
ibm-policyInstanceName:WeekdayPrime-time
ibm-ptpConditionTime:20000701000000:20050630235959
ibm-ptpConditionMonthOfYearMask:111111111111
ibm-ptpConditionDayOfMonthMask:11111111111111111111111111111111
ibm-ptpConditionDayOfWeekMask:0111110
ibm-ptpConditionTimeOfDayMask:060000:220000
ibm-ptpConditionLocalOrUtcTime:1
ibm-policyKeywords:POLICY
description:Active weekdays 6am - 10pm local time, 7/1/2000 to 7/1/2005
```

The goal of this policy is to ensure that outgoing data that match the specified attributes will be assigned a QoS service level defined in action "interactive1".

The following statements apply to the example in this topic:

- This rule will match only traffic on TCP connections (protocol 6) with a source port of 80 (i.e. HTTP server) and application defined data beginning with the string "/catalog".
- Because we are dealing with HTTP traffic, this rule is basically indicating that all outgoing traffic associated with a URI that begins with "/catalog" should be managed using the DS characteristics specified in the "interactive1" policy action.

```
dn:cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:web-catalog-rule
ibm-policyRuleName:web-catalog-rule
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListType:1
ibm-policyRulePriority:10
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Web catalog rule

dn:cn=condassoc1, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:web-catalog-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents first reusable DNF condition - TCP

dn:cn=condassoc2, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:web-catalog-condition2
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=webPort, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents second reusable DNF condition - web port

dn:cn=condassoc3, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:condassoc3
ibm-policyConditionName:web-catalog-condition3
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-applicationData:/catalog
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
```

```
description:Rule-specific condition - web catalog pages

dn:cn=actassoc1, cn=web-catalog-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:web-catalog-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=interactive1, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - interactive 1

dn: cn=interactive1, cn=QoSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:interactive1
ibm-policyActionName:interactive1-action
ibm-policyScope:DataTraffic
ibm-outgoingTOS:10000000
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS interactive 1 action (TOS 100)

dn:cn=webPort, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:webPort
ibm-policyConditionName:webPort-condition
ibm-sourcePortRange:80
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS web port condition

dn:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP-condition
ibm-protocolNumberRange:6
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS protocol TCP condition
```

RSVP policy example

The goal of this RSVP policy is to establish limits on resource reservations requested by RSVP applications using the RSVP API (RAPI) interface. The policy is identified as an RSVP policy by the ibm-PolicyScope:RSVP attribute in the ibm-PolicyActionInstance object.

The following statements apply to the example in this topic:

- The policy rule selects traffic from source ports in the range 8000 to 8001, with a protocol ID of 6 (TCP).
- One policy action specifies that the ToS byte be set to 01100000 for traffic that conforms to this policy.
- The RSVP policy action limits the type of RSVP service requested by RSVP applications to Controlled Load. Applications requesting Guaranteed service are downgraded to using Controlled Load service. In addition, the action limits the mean rate and token bucket size to 50000 bytes per second and 6000 bytes, respectively. These values are requested by RSVP applications in the traffic specification, or Tspec.
- The action also limits the number of active RSVP flows that map to this policy to 10.

```
dn:cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:intserv-rule
ibm-policyRuleName:intserv-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleConditionListDN:cn=condassoc1, cn=intserv-rule, cn=QoS, cn=advanced,
     ou=policy, o=IBM, c=US
ibm-policyRuleActionListDN:cn=actassoc1, cn=intserv-rule, cn=QoS, cn=advanced,
```

```
     ou=policy, o=IBM, c=US
ibm-policyRuleActionListDN:cn=actassoc2, cn=intserv-rule, cn=QoS, cn=advanced,
     ou=policy, o=IBM, c=US
ibm-policyRuleValidityPeriodList:cn=period2, cn=time, cn=repository, o=IBM, c=US
ibm-policyRuleValidityPeriodList:cn=period3, cn=time, cn=repository, o=IBM, c=US
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Integrated Services rule

dn:cn=condassoc1, cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-hostConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:condassoc1
ibm-policyConditionName:intserv-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-ProtocolNumberRange:6
ibm-SourceIPAddressRange:1
ibm-SourcePortRange:8000-8001
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific condition - all local IP addresses, application TCP ports

dn:cn=actassoc1, cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:actassoc1
ibm-policyActionName:intserv-action1
ibm-policyActionOrder:1
ibm-PolicyScope:DataTraffic
ibm-OutgoingTOS:01100000
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific action - set TOS

dn:cn=actassoc2, cn=intserv-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:actassoc2
ibm-policyActionName:intserv-action2
ibm-policyActionOrder:2
ibm-PolicyScope:RSVP
ibm-OutgoingTOS:01100000
ibm-FlowServiceType:ControlledLoad
ibm-MaxRatePerFlow:400
ibm-MaxTokenBucketPerFlow:48
ibm-MaxFlows:10
ibm-policyKeywords:Intserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific action - RSVP limitations

dn:cn=period2, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period2
ibm-policyConditionName:EndOfMonth-time
ibm-ptpConditionDayOfMonthMask:00000000000000000000000000000001
     00000000000000000000000000000000
ibm-ptpConditionLocalOrUtcTime:2
ibm-policyKeywords:POLICY
description:Active last day of the month (in UTC)

dn:cn=period3, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period3
ibm-policyConditionName:PacificNight-time
ibm-ptpConditionTimeOfDayMask:190000:030000
ibm-ptpConditionTimeZone:-08
ibm-policyKeywords:POLICY
description:Active 7pm - 3am local time, Pacific Time Zone (no daylight savings)
```

Sysplex distributor routing policy example

The goal of this sysplex distributor policy is to limit the number of SD target servers for inbound Telnet traffic. The policies are identified as SD policies by the ibm-policyGroupForLoadDistribution:TRUE attribute in the ibm-PolicyGroup object.

The following statements apply to the example in this topic:

- Separate policies are defined on the sysplex distributor distributing and target servers.
- The policy rules select incoming Telnet connection requests.
- The selected target server will be based on WLM information and QoS information if activated at the target servers.
- The rule (disttelnet) is coded on the distributing stack to select inbound traffic destined to the Telnet server.
- The rule (targtelnet) is coded on the target server to select outbound data from the Telnet server.
- If none of the specified target servers are available to service incoming requests (either the node is down or the Telnet server is not active), then sysplex distributor will distribute the requests to any available target server.

Note: If the ibm-Interface:1–0.0.0.0 attribute were not present, and none of the defined target servers were available, sysplex distributor would reject the request.

```
dn:cn=sysplex, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyGroup
objectclass:ibm-policyRuleContainmentAuxClass
objectclass:ibm-policyGroupLoadDistributionAuxClass
cn:sysplex
ibm-policyGroupName:QoSadvancedsysplex-Group
ibm-policyRulesAuxContainedSet:cn=disttelnet-rule,cn=QoS,cn=advanced,ou=policy,
     o=IBM,c=US
ibm-policyGroupForLoadDistribution:TRUE
ibm-policyKeywords:Sysplex
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description: QoS advanced examples sysplex group.

dn:cn=disttelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:disttelnet-rule
ibm-policyRuleName:disttelnet-rule
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:20
ibm-policyKeywords:Sysplex
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Sysplex Distributor telnet rule

dn:cn=condassoc1, cn=disttelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:condassoc1
ibm-policyConditionName:disttelnet-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-ProtocolNumberRange:6
ibm-DestinationPortRange:23
ibm-policyKeywords:Sysplex
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Rule-specific condition - telnet inbound SD traffic

dn:cn=actassoc1, cn=disttelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:disttelnet-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=telnetGold, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Sysplex
```

```
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - telnet Gold Service

dn:cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:targtelnet-rule
ibm-policyRuleName:targtelnet-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:20
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:QOS Sysplex Target telnet rule

dn:cn=condassoc1, cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:targtelnet-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable condition at level 1 - TCP

dn:cn=condassoc2, cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:targtelnet-condition2
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=telnetdPort, cn=QoScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable condition at level 2 - telnetd port

dn:cn=actassoc1, cn=targtelnet-rule, cn=QoS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:targtelnet-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=telnetGold, cn=QoSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Represents reusable action - telnet Gold Service

dn: cn=telnetGold, cn=QoSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-policyActionAuxClass
objectclass:ibm-serviceCategoriesAuxClass
cn:telnetGold
ibm-policyActionName:telnetGold-action
ibm-PolicyScope:DataTraffic
ibm-OutgoingTOS:10100000
ibm-MinRate:500
ibm-Interface:1--129.100.11.1
ibm-Interface:1--129.100.21.1
ibm-Interface:1--129.200.12.1
ibm-Interface:1--0.0.0.0
ibm-policyKeywords:Diffserv
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS telnet Gold Service action

dn:cn=telnetdPort, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:telnetdPort
ibm-policyConditionName:telnetdPort-condition
ibm-sourcePortRange:23
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS telnetd port condition
```

```
dn:cn=IpProtTCP, cn=QoScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-policyConditionAuxClass
objectclass:ibm-networkingPolicyConditionAuxClass
objectclass:ibm-applicationConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP-condition
ibm-protocolNumberRange:6
ibm-policyKeywords:QOS
ibm-policyKeywords:POLICY
description:Reusable QoS protocol TCP condition

dn:cn=period1, cn=time, cn=repository, o=IBM, c=US
objectclass:ibm-policyInstance
objectclass:ibm-policyTimePeriodConditionAuxClass
cn:period1
ibm-policyInstanceName:WeekdayPrime-time
ibm-ptpConditionTime:20000701000000:20050630235959
ibm-ptpConditionMonthOfYearMask:111111111111
ibm-ptpConditionDayOfMonthMask:11111111111111111111111111111111
ibm-ptpConditionDayOfWeekMask:0111110
ibm-ptpConditionTimeOfDayMask:060000:220000
ibm-ptpConditionLocalOrUtcTime:1
ibm-policyKeywords:POLICY
description:Active weekdays 6am - 10pm local time, 7/1/2000 to 7/1/2005
```

Defining IDS policies using LDAP

To define IDS policies using LDAP, see the following examples.

IDS scan policy example

```
dn:cn=scanglobal-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:scanglobal-rule
ibm-policyRuleName:ScanGlobal-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS global scan rule

dn:cn=condassoc1, cn=scanglobal-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsScanConditionAuxClass
cn:condassoc1
ibm-policyConditionName:ScanGlobal-condition
ibm-idsConditionType:SCAN_GLOBAL
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition

dn:cn=actassoc1, cn=scanglobal-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsScanActionAuxClass
objectclass:ibm-idsNotificationAuxClass
cn:actassoc1
ibm-policyActionName:ScanGlobal-action
ibm-idsActionType:SCAN_GLOBAL
ibm-policyActionOrder:1
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsNotification:SYSLOGDETAIL
ibm-idsLoggingLevel:4
ibm-idsTraceData:RECORDSIZE
```

```
ibm-idsTraceRecordSize:200
ibm-idsFSInterval:2
ibm-idsFSThreshold:5
ibm-idsSSInterval:480
ibm-idsSSThreshold:10
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - Fast scan = 5 in 2 minutes, Slow scan = 10 in 8
hours

dn:cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:scaneventlow-rule
ibm-policyRuleName:ScanEventLow-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS scan event rule for low sensitivity on TCP and UDP Low Ports

dn:cn=condassoc2, cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc2
ibm-policyConditionName:ScanEventLow-condition2
ibm-policyConditionDN:cn=ScanTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable CNF condition level 2 - scan TCP low ports

dn:cn=condassoc3, cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc3
ibm-policyConditionName:ScanEventLow-condition3
ibm-policyConditionDN:cn=ScanUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable CNF condition level 2 - scan UDP low ports

dn:cn=actassoc1, cn=scaneventlow-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsScanSensitivityActionAuxClass
objectclass:ibm-idsScanExclusionActionAuxClass
cn:actassoc1
ibm-policyActionName:ScanEventLow-action
ibm-idsActionType:SCAN_EVENT
ibm-policyActionOrder:1
ibm-idsSensitivity:LOW
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - low sensitivity

dn:cn=scaneventmedium-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:scaneventmedium-rule
ibm-policyRuleName:ScanEventMedium-rule
ibm-policyRuleConditionListType:2
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
```

```
description:IDS scan event rule for medium sensitivity on ICMP

dn:cn=condassoc1, cn=scaneventmedium-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsScanEventConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:condassoc1
ibm-policyConditionName:ScanEventMedium-condition
ibm-idsConditionType:SCAN_EVENT
ibm-policyConditionGroupNumber:2
ibm-policyConditionNegated:FALSE
ibm-idsProtocolRange:1
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - ICMP protocol

dn:cn=actassoc1, cn=scaneventmedium-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsScanSensitivityActionAuxClass
cn:actassoc1
ibm-policyActionName:ScanEventMedium-action
ibm-idsActionType:SCAN_EVENT
ibm-policyActionOrder:1
ibm-idsSensitivity:MEDIUM
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - medium sensitivity

dn:cn=ScanTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:SCAN_EVENT
objectclass:ibm-idsScanEventConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:ScanTcpLowPorts
ibm-policyConditionName:ScanTcpLowPorts-condition
ibm-idsProtocolRange:6
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS Scan TCP Low Ports condition

dn:cn=ScanUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:SCAN_EVENT
objectclass:ibm-idsScanEventConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:ScabUdpLowPorts
ibm-policyConditionName:ScanUdpLowPorts-condition
ibm-idsProtocolRange:17
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:Scan
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS Scan UDP Low Ports condition
```

IDS attack policy example

```
dn:cn=attackMalformed-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackMalformed-rule
ibm-policyRuleName:AttackMalformed-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
```

```
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for Malformed Packets

dn:cn=condassoc1, cn=attackMalformed-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackMalformed-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:MALFORMED_PACKET
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackMalformed-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:attackMalformed-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackFlood-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackFlood-rule
ibm-policyRuleName:AttackFlood-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for Floods

dn:cn=condassoc1, cn=attackFlood-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackFlood-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:FLOOD
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackFlood-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsAttackActionsAuxClass
objectclass:ibm-idsFloodAttackActionsAuxClass
cn:actassoc1
ibm-policyActionName:attackFlood-action
ibm-policyActionOrder:1
ibm-idsActionType:ATTACK
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:1
ibm-idsTypeActions:EXCEPTSTATS
```

```
ibm-idsStatInterval:60
ibm-idsTraceData:RECORDSIZE
ibm-idsTraceRecordSize:200
ibm-idsIfcFloodPercentage:10
ibm-idsIfcFloodMinDiscard:1000
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific action - LOG(SYSLOG(1) NOCONSOLE) EXCEPTSTATS(60)
TRACE(200)
dn:cn=attackICMPRedirect-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackICMPRedirect-rule
ibm-policyRuleName:AttackICMPRedirect-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for ICMP Redirect

dn:cn=condassoc1, cn=attackICMPRedirect-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackICMPRedirect-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
ibm-idsAttackType:ICMP_REDIRECT
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackICMPRedirect-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:attackICMPRedirect-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackIpFragment-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackIpFragment-rule
ibm-policyRuleName:AttackIpFragment-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for IP fragment restriction

dn:cn=condassoc1, cn=attackIpFragment-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:condassoc1
ibm-policyConditionName:attackIpFragment-condition
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-idsConditionType:ATTACK
```

```
ibm-idsAttackType:IP_FRAGMENT
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Rule-specific condition - attack type

dn:cn=actassoc1, cn=attackIpFragment-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:attackIpFragment-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackIpProt-rule
ibm-policyRuleName:AttackIPprot-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for restricted protocol

dn:cn=condassoc1, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:AttackIPprot-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=attackIpProtcond1, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents first reusable DNF condition at level 1

dn:cn=condassoc1a, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1a
ibm-policyConditionName:AttackIPprot-condition1a
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtICMP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents second reusable DNF condition at level 1 (negated) allow ICMP

dn:cn=condassoc1b, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1b
ibm-policyConditionName:AttackIPprot-condition1b
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtTCP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents third reusable DNF condition at level 1 (negated) allow TCP

dn:cn=condassoc1c, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1c
ibm-policyConditionName:AttackIPprot-condition1c
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtUDP, cn=IDScond, cn=repository, o=IBM, c=US
```

```
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents fourth reusable DNF condition at level 1 (negated) allow UDP

dn:cn=actassoc1, cn=attackIpProt-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:AttackIPprot-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn:cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:attackOutboundRaw-rule
ibm-policyRuleName:AttackOutboundRaw-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRulePriority:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS attack rule for Outbound Raw restrictions

dn:cn=condassoc1, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:AttackOutboundRaw-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=attackOutboundRawcond1, cn=IDScond, cn=repository, o=IBM,
c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents first reusable DNF condition at level 1

dn:cn=condassoc1a, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1a
ibm-policyConditionName:AttackOutboundRaw-condition1a
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtICMP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents second reusable DNF condition at level 1 (negated) allow ICMP

dn:cn=condassoc1b, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1b
ibm-policyConditionName:AttackOutboundRaw-condition1b
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtUDP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents third reusable DNF condition at level 1 (negated) allow UDP

dn:cn=condassoc1c, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1c
```

```
ibm-policyConditionName:AttackOutboundRaw-condition1c
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtIGMP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents fourth reusable DNF condition at level 1 (negated) allow IGMP

dn:cn=condassoc1d, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1d
ibm-policyConditionName:AttackOutboundRaw-condition1d
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:TRUE
ibm-policyConditionDN:cn=IpProtOSPFIGP, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents fifth reusable DNF condition at level 1 (negated) allow
OSPFIGP

dn:cn=actassoc1, cn=attackOutboundRaw-rule, cn=IDS, cn=starter, ou=policy, o=IBM,
c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:AttackOutboundRaw-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - attack action 1

dn: cn=attackact1, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsAttackActionsAuxClass
cn:attackact1
ibm-policyActionName:AttackLog-action
ibm-idsActionType:ATTACK
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:1
ibm-idsTypeActions:EXCEPTSTATS
ibm-idsStatInterval:60
ibm-idsTraceData:RECORDSIZE
ibm-idsTraceRecordSize:200
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS common attack action - LOG(SYSLOG(1) NOCONSOLE) NOLIMIT
description:IDS common attack action - EXECPTSTATS(60) TRACE(200)

dn:cn=attackIpProtcond1, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsAttackConditionAuxClass
cn:attackIpProtcond1
ibm-policyConditionName:AttackIPprot-condition1
ibm-idsConditionType:ATTACK
ibm-idsAttackType:RESTRICTED_IP_PROTOCOL
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS attack condition 1 for restricted IP protocol

dn:cn=attackOutboundRawcond1, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
```

```
objectclass:ibm-idsAttackConditionAuxClass
cn:attackOutboundRawcond1
ibm-policyConditionName:AttackOutboundRaw-condition1
ibm-idsConditionType:ATTACK
ibm-idsAttackType:OUTBOUND_RAW
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS attack condition 1 for Outbound Raw restrictions

dn:cn=IpProtICMP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtICMP
ibm-policyConditionName:IpProtICMP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:1
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol ICMP

dn:cn=IpProtIGMP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtIGMP
ibm-policyConditionName:IpProtIGMP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:2
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol IGMP

dn:cn=IpProtTCP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtTCP
ibm-policyConditionName:IpProtTCP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:6
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol TCP

dn:cn=IpProtUDP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtUDP
ibm-policyConditionName:IpProtUDP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:17
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol UDP

dn:cn=IpProtOSPFIGP, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:IpProtOSPFIGP
ibm-policyConditionName:IpProtOSPFIGP
ibm-idsConditionType:ATTACK
ibm-idsProtocolRange:89
ibm-policyKeywords:Attack
ibm-policyKeywords:IDS
```

```
ibm-policyKeywords:POLICY
description:Reusable IDS condition for IP protocol OSPFIGP
```

IDS TCP traffic regulation policy example

```
dn:cn=trtcp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:trtcp-rule
ibm-policyRuleName:TRtcp-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListDN:cn=condassoc1,cn=trtcp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRuleActionListDN:cn=actassoc1,cn=trtcp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRulePriority:2
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS TR TCP rule

dn:cn=condassoc1, cn=trtcp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:TRtcp-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=TrTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable condition - TR TCP low ports

dn:cn=actassoc1, cn=trtcp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:TRtcp-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=trtcpact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - TR TCP action 1

dn:cn=trtcpWeb-rule, cn=IDS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:trtcpWeb-rule
ibm-policyRuleName:trtcpWeb-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListDN:cn=condassoc1,cn=trtcpWeb-rule,cn=IDS,cn=advanced,
     ou=policy, o=IBM,c=US
ibm-policyRuleActionListDN:cn=actassoc1,cn=trtcpWeb-rule,cn=IDS,cn=advanced,
     ou=policy,o=IBM,c=US
ibm-policyRuleValidityPeriodList:cn=period1, cn=time, cn=repository, o=IBM, c=US
ibm-policyRulePriority:7
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS TR TCP rule with limit

dn:cn=condassoc1, cn=trtcpWeb-rule, cn=IDS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:TRtcpWeb-condition1
ibm-policyConditionGroupNumber:1
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=TrTcpWebPort, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable condition - TR TCP web port

dn:cn=actassoc1, cn=trtcpWeb-rule, cn=IDS, cn=advanced, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:TRtcpWeb-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=trtcpact2, cn=IDSact, cn=repository, o=IBM, c=US
```

```
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - TR TCP action 2

dn: cn=trtcpact1, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsTRtcpActionAuxClass
cn:trtcpact1
ibm-policyActionName:TRtcpLog-action
ibm-idsActionType:TR
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:4
ibm-idsTypeActions:STATISTICS
ibm-idsStatInterval:60
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS TR TCP action TCP(64K,100%) LOG(SYSLOG(4) NOCONSOLE) NOLIMIT
description:TRACE(HEADER) STATISTICS(60)

dn:cn=trtcpact2, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsTRtcpActionAuxClass
cn:trtcpact2
ibm-policyActionName:TRtcpLimit-action
ibm-idsActionType:TR
ibm-idsTypeActions:LIMIT
ibm-idsTypeActions:LOG
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:4
ibm-idsTypeActions:EXCEPTSTATS
ibm-idsStatInterval:60
ibm-idsTRtcpTotalConnections:1000
ibm-idsTRtcpPercentage:10
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS TR TCP action TCP(1K,10%) LOG(SYSLOG(4) NOCONSOLE) LIMIT
description:TRACE(HEADER) EXCEPTSTATS(60)

dn:cn=TrTcpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:TR
objectclass:ibm-idsTrafficRegulationConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:TrTcpLowPorts
ibm-policyConditionName:TrTcpLowPorts-condition
ibm-idsProtocolRange:6
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS TR TCP Low Ports condition

dn:cn=TrTcpWebPort, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:TR
objectclass:ibm-idsTrafficRegulationConditionAuxClass
objectclass:ibm-idsHostConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:TrTcpWebPort
ibm-policyConditionName:TrTcpWebPort-condition
ibm-idsProtocolRange:6
ibm-idsLocalPortRange:80
ibm-idsLocalHostIPAddress:3-10.14.243.87
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS TR TCP Web Port condition
```

IDS UDP traffic regulation policy example

```
dn:cn=trudp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRule
cn:trudp-rule
ibm-policyRuleName:TRudp-rule
ibm-policyRuleConditionListType:1
ibm-policyRuleEnabled:1
ibm-policyRuleConditionListDN:cn=condassoc1,cn=trudp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRuleActionListDN:cn=actassoc1,cn=trudp-rule,cn=IDS,cn=starter,
     ou=policy,o=IBM,c=US
ibm-policyRulePriority:2
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Example of IDS TR UDP rule

dn:cn=condassoc1, cn=trudp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleConditionAssociation
cn:condassoc1
ibm-policyConditionName:TRudp-condition1
ibm-policyConditionGroupNumber:7
ibm-policyConditionNegated:FALSE
ibm-policyConditionDN:cn=TrUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable condition - TR UDP low ports

dn:cn=actassoc1, cn=trudp-rule, cn=IDS, cn=starter, ou=policy, o=IBM, c=US
objectclass:ibm-policyRuleActionAssociation
cn:actassoc1
ibm-policyActionName:TRudp-action
ibm-policyActionOrder:1
ibm-policyActionDN:cn=trudpact1, cn=IDSact, cn=repository, o=IBM, c=US
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Represents reusable action - TR UDP action 1

dn: cn=trudpact1, cn=IDSact, cn=repository, o=IBM, c=US
objectclass:ibm-policyActionInstance
objectclass:ibm-idsActionAuxClass
objectclass:ibm-idsNotificationAuxClass
objectclass:ibm-idsTRudpActionAuxClass
cn:trudpact1
ibm-policyActionName:TRudpLog-action
ibm-idsActionType:TR
ibm-idsTypeActions:LOG
ibm-idsTypeActions:LIMIT
ibm-idsNotification:SYSLOG
ibm-idsLoggingLevel:4
ibm-idsTypeActions:STATISTICS
ibm-idsStatInterval:60
ibm-idsTRudpQueueSize:LONG
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:IDS TR UDP action UDPQ(LONG) LOG(SYSLOG(4) NOCONSOLE) LIMIT
description:TRACE(HEADER) STATISTICS(60)
dn:cn=TrUdpLowPorts, cn=IDScond, cn=repository, o=IBM, c=US
objectclass:ibm-policyConditionInstance
objectclass:ibm-idsConditionAuxClass
ibm-idsConditionType:TR
objectclass:ibm-idsTrafficRegulationConditionAuxClass
objectclass:ibm-idsTransportConditionAuxClass
cn:TrUdpLowPorts
ibm-policyConditionName:TrUdpLowPorts-condition
ibm-idsProtocolRange:17
ibm-idsLocalPortRange:1-1023
ibm-policyKeywords:TR
ibm-policyKeywords:IDS
ibm-policyKeywords:POLICY
description:Reusable IDS TR UDP Low Ports condition
```

Appendix G. Related protocol specifications

This appendix lists the related protocol specifications (RFCs) for TCP/IP. The Internet Protocol suite is still evolving through requests for comments (RFC). New protocols are being designed and implemented by researchers and are brought to the attention of the Internet community in the form of RFCs. Some of these protocols are so useful that they become recommended protocols. That is, all future implementations for TCP/IP are recommended to implement these particular functions or protocols. These become the *de facto* standards, on which the TCP/IP protocol suite is built.

RFCs are available at http://www.rfc-editor.org/rfc.html.

Draft RFCs that have been implemented in this and previous Communications Server releases are listed at the end of this topic.

Many features of TCP/IP Services are based on the following RFCs:

RFC
Title and Author

RFC 652
Telnet output carriage-return disposition option D. Crocker

RFC 653
Telnet output horizontal tabstops option D. Crocker

RFC 654
Telnet output horizontal tab disposition option D. Crocker

RFC 655
Telnet output formfeed disposition option D. Crocker

RFC 657
Telnet output vertical tab disposition option D. Crocker

RFC 658
Telnet output linefeed disposition D. Crocker

RFC 698
Telnet extended ASCII option T. Mock

RFC 726
Remote Controlled Transmission and Echoing Telnet option J. Postel, D. Crocker

RFC 727
Telnet logout option M.R. Crispin

RFC 732
Telnet Data Entry Terminal option J.D. Day

RFC 733
Standard for the format of ARPA network text messages D. Crocker, J. Vittal, K.T. Pogran, D.A. Henderson

RFC 734
SUPDUP Protocol M.R. Crispin

RFC 735
Revised Telnet byte macro option D. Crocker, R.H. Gumpertz

RFC 736
Telnet SUPDUP option M.R. Crispin

RFC 749
Telnet SUPDUP—Output option B. Greenberg

RFC 765
File Transfer Protocol specification J. Postel

© Copyright IBM Corp. 2000, 2022

RFC 768
User Datagram Protocol J. Postel

RFC 779
Telnet send-location option E. Killian

RFC 791
Internet Protocol J. Postel

RFC 792
Internet Control Message Protocol J. Postel

RFC 793
Transmission Control Protocol J. Postel

RFC 820
Assigned numbers J. Postel

RFC 823
DARPA Internet gateway R. Hinden, A. Sheltzer

RFC 826
Ethernet Address Resolution Protocol: Or converting network protocol addresses to 48.bit Ethernet address for transmission on Ethernet hardware D. Plummer

RFC 854
Telnet Protocol Specification J. Postel, J. Reynolds

RFC 855
Telnet Option Specification J. Postel, J. Reynolds

RFC 856
Telnet Binary Transmission J. Postel, J. Reynolds

RFC 857
Telnet Echo Option J. Postel, J. Reynolds

RFC 858
Telnet Suppress Go Ahead Option J. Postel, J. Reynolds

RFC 859
Telnet Status Option J. Postel, J. Reynolds

RFC 860
Telnet Timing Mark Option J. Postel, J. Reynolds

RFC 861
Telnet Extended Options: List Option J. Postel, J. Reynolds

RFC 862
Echo Protocol J. Postel

RFC 863
Discard Protocol J. Postel

RFC 864
Character Generator Protocol J. Postel

RFC 865
Quote of the Day Protocol J. Postel

RFC 868
Time Protocol J. Postel, K. Harrenstien

RFC 877
Standard for the transmission of IP datagrams over public data networks J.T. Korb

RFC 883
Domain names: Implementation specification P.V. Mockapetris

RFC 884
Telnet terminal type option M. Solomon, E. Wimmers

RFC 885
Telnet end of record option J. Postel

RFC 894
Standard for the transmission of IP datagrams over Ethernet networks C. Hornig

RFC 896
Congestion control in IP/TCP internetworks J. Nagle

RFC 903
Reverse Address Resolution Protocol R. Finlayson, T. Mann, J. Mogul, M. Theimer

RFC 904
Exterior Gateway Protocol formal specification D. Mills

RFC 919
Broadcasting Internet Datagrams J. Mogul

RFC 922
Broadcasting Internet datagrams in the presence of subnets J. Mogul

RFC 927
TACACS user identification Telnet option B.A. Anderson

RFC 933
Output marking Telnet option S. Silverman

RFC 946
Telnet terminal location number option R. Nedved

RFC 950
Internet Standard Subnetting Procedure J. Mogul, J. Postel

RFC 952
DoD Internet host table specification K. Harrenstien, M. Stahl, E. Feinler

RFC 959
File Transfer Protocol J. Postel, J.K. Reynolds

RFC 961
Official ARPA-Internet protocols J.K. Reynolds, J. Postel

RFC 974
Mail routing and the domain system C. Partridge

RFC 1001
Protocol standard for a NetBIOS service on a TCP/UDP transport: Concepts and methods NetBios Working Group in the Defense Advanced Research Projects Agency, Internet Activities Board, End-to-End Services Task Force

RFC 1002
Protocol Standard for a NetBIOS service on a TCP/UDP transport: Detailed specifications NetBios Working Group in the Defense Advanced Research Projects Agency, Internet Activities Board, End-to-End Services Task Force

RFC 1006
ISO transport services on top of the TCP: Version 3 M.T. Rose, D.E. Cass

RFC 1009
Requirements for Internet gateways R. Braden, J. Postel

RFC 1011
Official Internet protocols J. Reynolds, J. Postel

RFC 1013
X Window System Protocol, version 11: Alpha update April 1987 R. Scheifler

RFC 1014
XDR: External Data Representation standard Sun Microsystems

RFC 1027
Using ARP to implement transparent subnet gateways S. Carl-Mitchell, J. Quarterman

RFC 1032
Domain administrators guide M. Stahl

RFC 1033
Domain administrators operations guide M. Lottor

RFC 1034
Domain names—concepts and facilities P.V. Mockapetris

RFC 1035
Domain names—implementation and specification P.V. Mockapetris

RFC 1038
Draft revised IP security option M. St. Johns

RFC 1041
Telnet 3270 regime option Y. Rekhter

RFC 1042
Standard for the transmission of IP datagrams over IEEE 802 networks J. Postel, J. Reynolds

RFC 1043
Telnet Data Entry Terminal option: DODIIS implementation A. Yasuda, T. Thompson

RFC 1044
Internet Protocol on Network System's HYPERchannel: Protocol specification K. Hardwick, J. Lekashman

RFC 1053
Telnet X.3 PAD option S. Levy, T. Jacobson

RFC 1055
Nonstandard for transmission of IP datagrams over serial lines: SLIP J. Romkey

RFC 1057
RPC: Remote Procedure Call Protocol Specification: Version 2 Sun Microsystems

RFC 1058
Routing Information Protocol C. Hedrick

RFC 1060
Assigned numbers J. Reynolds, J. Postel

RFC 1067
Simple Network Management Protocol J.D. Case, M. Fedor, M.L. Schoffstall, J. Davin

RFC 1071
Computing the Internet checksum R.T. Braden, D.A. Borman, C. Partridge

RFC 1072
TCP extensions for long-delay paths V. Jacobson, R.T. Braden

RFC 1073
Telnet window size option D. Waitzman

RFC 1079
Telnet terminal speed option C. Hedrick

RFC 1085
ISO presentation services on top of TCP/IP based internets M.T. Rose

RFC 1091
Telnet terminal-type option J. VanBokkelen

RFC 1094
NFS: Network File System Protocol specification Sun Microsystems

RFC 1096
Telnet X display location option G. Marcy

RFC 1101
DNS encoding of network names and other types P. Mockapetris

RFC 1112
Host extensions for IP multicasting S.E. Deering

RFC 1113
Privacy enhancement for Internet electronic mail: Part I — message encipherment and authentication procedures J. Linn

RFC 1118
Hitchhikers Guide to the Internet E. Krol

RFC 1122
Requirements for Internet Hosts—Communication Layers R. Braden, Ed.

RFC 1123
Requirements for Internet Hosts—Application and Support R. Braden, Ed.

RFC 1146
TCP alternate checksum options J. Zweig, C. Partridge

RFC 1155
Structure and identification of management information for TCP/IP-based internets M. Rose, K. McCloghrie

RFC 1156
Management Information Base for network management of TCP/IP-based internets K. McCloghrie, M. Rose

RFC 1157
Simple Network Management Protocol (SNMP) J. Case, M. Fedor, M. Schoffstall, J. Davin

RFC 1158
Management Information Base for network management of TCP/IP-based internets: MIB-II M. Rose

RFC 1166
Internet numbers S. Kirkpatrick, M.K. Stahl, M. Recker

RFC 1179
Line printer daemon protocol L. McLaughlin

RFC 1180
TCP/IP tutorial T. Socolofsky, C. Kale

RFC 1183
New DNS RR Definitions C.F. Everhart, L.A. Mamakos, R. Ullmann, P.V. Mockapetris

RFC 1184
Telnet Linemode Option D. Borman

RFC 1186
MD4 Message Digest Algorithm R.L. Rivest

RFC 1187
Bulk Table Retrieval with the SNMP M. Rose, K. McCloghrie, J. Davin

RFC 1188
Proposed Standard for the Transmission of IP Datagrams over FDDI Networks D. Katz

RFC 1190
Experimental Internet Stream Protocol: Version 2 (ST-II) C. Topolcic

RFC 1191
Path MTU discovery J. Mogul, S. Deering

RFC 1198
FYI on the X window system R. Scheifler

RFC 1207
FYI on Questions and Answers: Answers to commonly asked "experienced Internet user" questions G. Malkin, A. Marine, J. Reynolds

RFC 1208
Glossary of networking terms O. Jacobsen, D. Lynch

RFC 1213
Management Information Base for Network Management of TCP/IP-based internets: MIB-II K. McCloghrie, M.T. Rose

RFC 1215
Convention for defining traps for use with the SNMP M. Rose

RFC 1227
SNMP MUX protocol and MIB M.T. Rose

RFC 1228
SNMP-DPI: Simple Network Management Protocol Distributed Program Interface G. Carpenter, B. Wijnen

RFC 1229
Extensions to the generic-interface MIB K. McCloghrie

RFC 1230
IEEE 802.4 Token Bus MIB K. McCloghrie, R. Fox

RFC 1231
IEEE 802.5 Token Ring MIB K. McCloghrie, R. Fox, E. Decker

RFC 1236
IP to X.121 address mapping for DDN L. Morales, P. Hasse

RFC 1256
ICMP Router Discovery Messages S. Deering, Ed.

RFC 1267
Border Gateway Protocol 3 (BGP-3) K. Lougheed, Y. Rekhter

RFC 1268
Application of the Border Gateway Protocol in the Internet Y. Rekhter, P. Gross

RFC 1269
Definitions of Managed Objects for the Border Gateway Protocol: Version 3 S. Willis, J. Burruss

RFC 1270
SNMP Communications Services F. Kastenholz, ed.

RFC 1285
FDDI Management Information Base J. Case

RFC 1315
Management Information Base for Frame Relay DTEs C. Brown, F. Baker, C. Carvalho

RFC 1321
The MD5 Message-Digest Algorithm R. Rivest

RFC 1323
TCP Extensions for High Performance V. Jacobson, R. Braden, D. Borman

RFC 1325
FYI on Questions and Answers: Answers to Commonly Asked "New Internet User" Questions G. Malkin, A. Marine

RFC 1327
Mapping between X.400 (1988)/ISO 10021 and RFC 822 S. Hardcastle-Kille

RFC 1340
Assigned Numbers J. Reynolds, J. Postel

RFC 1344
Implications of MIME for Internet Mail Gateways N. Bornstein

RFC 1349
Type of Service in the Internet Protocol Suite P. Almquist

RFC 1351
SNMP Administrative Model J. Davin, J. Galvin, K. McCloghrie

RFC 1352
SNMP Security Protocols J. Galvin, K. McCloghrie, J. Davin

RFC 1353
Definitions of Managed Objects for Administration of SNMP Parties K. McCloghrie, J. Davin, J. Galvin

RFC 1354
IP Forwarding Table MIB F. Baker

RFC 1356
Multiprotocol Interconnect® on X.25 and ISDN in the Packet Mode A. Malis, D. Robinson, R. Ullmann

RFC 1358
Charter of the Internet Architecture Board (IAB) L. Chapin

RFC 1363
A Proposed Flow Specification C. Partridge

RFC 1368
Definition of Managed Objects for IEEE 802.3 Repeater Devices D. McMaster, K. McCloghrie

RFC 1372
Telnet Remote Flow Control Option C. L. Hedrick, D. Borman

RFC 1374
IP and ARP on HIPPI J. Renwick, A. Nicholson

RFC 1381
SNMP MIB Extension for X.25 LAPB D. Throop, F. Baker

RFC 1382
SNMP MIB Extension for the X.25 Packet Layer D. Throop

RFC 1387
RIP Version 2 Protocol Analysis G. Malkin

RFC 1388
RIP Version 2 Carrying Additional Information G. Malkin

RFC 1389
RIP Version 2 MIB Extensions G. Malkin, F. Baker

RFC 1390
Transmission of IP and ARP over FDDI Networks D. Katz

RFC 1393
Traceroute Using an IP Option G. Malkin

RFC 1398
Definitions of Managed Objects for the Ethernet-Like Interface Types F. Kastenholz

RFC 1408
Telnet Environment Option D. Borman, Ed.

RFC 1413
Identification Protocol M. St. Johns

RFC 1416
Telnet Authentication Option D. Borman, ed.

RFC 1420
SNMP over IPX S. Bostock

RFC 1428
Transition of Internet Mail from Just-Send-8 to 8bit-SMTP/MIME G. Vaudreuil

RFC 1442
Structure of Management Information for version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1443
Textual Conventions for version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1445
Administrative Model for version 2 of the Simple Network Management Protocol (SNMPv2) J. Galvin, K. McCloghrie

RFC 1447
Party MIB for version 2 of the Simple Network Management Protocol (SNMPv2) K. McCloghrie, J. Galvin

RFC 1448
Protocol Operations for version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1464
Using the Domain Name System to Store Arbitrary String Attributes R. Rosenbaum

RFC 1469
IP Multicast over Token-Ring Local Area Networks T. Pusateri

RFC 1483
Multiprotocol Encapsulation over ATM Adaptation Layer 5 Juha Heinanen

RFC 1514
Host Resources MIB P. Grillo, S. Waldbusser

RFC 1516
Definitions of Managed Objects for IEEE 802.3 Repeater Devices D. McMaster, K. McCloghrie

RFC 1521
MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies N. Borenstein, N. Freed

RFC 1535
A Security Problem and Proposed Correction With Widely Deployed DNS Software E. Gavron

RFC 1536
Common DNS Implementation Errors and Suggested Fixes A. Kumar, J. Postel, C. Neuman, P. Danzig, S. Miller

RFC 1537
Common DNS Data File Configuration Errors P. Beertema

RFC 1540
Internet Official Protocol Standards J. Postel

RFC 1571
Telnet Environment Option Interoperability Issues D. Borman

RFC 1572
Telnet Environment Option S. Alexander

RFC 1573
Evolution of the Interfaces Group of MIB-II K. McCloghrie, F. Kastenholz

RFC 1577
Classical IP and ARP over ATM M. Laubach

RFC 1583
OSPF Version 2 J. Moy

RFC 1591
Domain Name System Structure and Delegation J. Postel

RFC 1592
Simple Network Management Protocol Distributed Protocol Interface Version 2.0 B. Wijnen, G. Carpenter, K. Curran, A. Sehgal, G. Waters

RFC 1594
FYI on Questions and Answers— Answers to Commonly Asked "New Internet User" Questions A. Marine, J. Reynolds, G. Malkin

RFC 1644
T/TCP — TCP Extensions for Transactions Functional Specification R. Braden

RFC 1646
TN3270 Extensions for LUname and Printer Selection C. Graves, T. Butts, M. Angel

RFC 1647
TN3270 Enhancements B. Kelly

RFC 1652
SMTP Service Extension for 8bit-MIMEtransport J. Klensin, N. Freed, M. Rose, E. Stefferud, D. Crocker

RFC 1664
Using the Internet DNS to Distribute RFC1327 Mail Address Mapping Tables C. Allochio, A. Bonito, B. Cole, S. Giordano, R. Hagens

RFC 1693
An Extension to TCP: Partial Order Service T. Connolly, P. Amer, P. Conrad

RFC 1695
Definitions of Managed Objects for ATM Management Version 8.0 using SMIv2 M. Ahmed, K. Tesink

RFC 1701
Generic Routing Encapsulation (GRE) S. Hanks, T. Li, D. Farinacci, P. Traina

RFC 1702
Generic Routing Encapsulation over IPv4 networks S. Hanks, T. Li, D. Farinacci, P. Traina

RFC 1706
DNS NSAP Resource Records B. Manning, R. Colella

RFC 1712
DNS Encoding of Geographical Location C. Farrell, M. Schulze, S. Pleitner D. Baldoni

RFC 1713
Tools for DNS debugging A. Romao

RFC 1723
RIP Version 2—Carrying Additional Information G. Malkin

RFC 1752
The Recommendation for the IP Next Generation Protocol S. Bradner, A. Mankin

RFC 1766
Tags for the Identification of Languages H. Alvestrand

RFC 1771
A Border Gateway Protocol 4 (BGP-4) Y. Rekhter, T. Li

RFC 1794
DNS Support for Load Balancing T. Brisco

RFC 1819
Internet Stream Protocol Version 2 (ST2) Protocol Specification—Version ST2+ L. Delgrossi, L. Berger Eds.

RFC 1826
IP Authentication Header R. Atkinson

RFC 1828
IP Authentication using Keyed MD5 P. Metzger, W. Simpson

RFC 1829
The ESP DES-CBC Transform P. Karn, P. Metzger, W. Simpson

RFC 1830
SMTP Service Extensions for Transmission of Large and Binary MIME Messages G. Vaudreuil

RFC 1831
RPC: Remote Procedure Call Protocol Specification Version 2 R. Srinivasan

RFC 1832
XDR: External Data Representation Standard R. Srinivasan

RFC 1833
Binding Protocols for ONC RPC Version 2 R. Srinivasan

RFC 1850
OSPF Version 2 Management Information Base F. Baker, R. Coltun

RFC 1854
SMTP Service Extension for Command Pipelining N. Freed

RFC 1869
SMTP Service Extensions J. Klensin, N. Freed, M. Rose, E. Stefferud, D. Crocker

RFC 1870
SMTP Service Extension for Message Size Declaration J. Klensin, N. Freed, K. Moore

RFC 1876
A Means for Expressing Location Information in the Domain Name System C. Davis, P. Vixie, T. Goodwin, I. Dickinson

RFC 1883
Internet Protocol, Version 6 (IPv6) Specification S. Deering, R. Hinden

RFC 1884
IP Version 6 Addressing Architecture R. Hinden, S. Deering, Eds.

RFC 1886
DNS Extensions to support IP version 6 S. Thomson, C. Huitema

RFC 1888
OSI NSAPs and IPv6 J. Bound, B. Carpenter, D. Harrington, J. Houldsworth, A. Lloyd

RFC 1891
SMTP Service Extension for Delivery Status Notifications K. Moore

RFC 1892
The Multipart/Report Content Type for the Reporting of Mail System Administrative Messages G. Vaudreuil

RFC 1894
*An Extensible Message Format for Delivery Status Notifications*K. Moore, G. Vaudreuil

RFC 1901
Introduction to Community-based SNMPv2 J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1902
Structure of Management Information for Version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1903
Textual Conventions for Version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1904
Conformance Statements for Version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1905
Protocol Operations for Version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1906
Transport Mappings for Version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1907
Management Information Base for Version 2 of the Simple Network Management Protocol (SNMPv2) J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1908
Coexistence between Version 1 and Version 2 of the Internet-standard Network Management Framework J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 1912
Common DNS Operational and Configuration Errors D. Barr

RFC 1918
Address Allocation for Private Internets Y. Rekhter, B. Moskowitz, D. Karrenberg, G.J. de Groot, E. Lear

RFC 1928
SOCKS Protocol Version 5 M. Leech, M. Ganis, Y. Lee, R. Kuris, D. Koblas, L. Jones

RFC 1930
Guidelines for creation, selection, and registration of an Autonomous System (AS) J. Hawkinson, T. Bates

RFC 1939
Post Office Protocol-Version 3 J. Myers, M. Rose

RFC 1981
Path MTU Discovery for IP version 6 J. McCann, S. Deering, J. Mogul

RFC 1982
Serial Number Arithmetic R. Elz, R. Bush

RFC 1985
SMTP Service Extension for Remote Message Queue Starting J. De Winter

RFC 1995
Incremental Zone Transfer in DNS M. Ohta

RFC 1996
A Mechanism for Prompt Notification of Zone Changes (DNS NOTIFY) P. Vixie

RFC 2010
Operational Criteria for Root Name Servers B. Manning, P. Vixie

RFC 2011
SNMPv2 Management Information Base for the Internet Protocol using SMIv2 K. McCloghrie, Ed.

RFC 2012
SNMPv2 Management Information Base for the Transmission Control Protocol using SMIv2 K. McCloghrie, Ed.

RFC 2013
SNMPv2 Management Information Base for the User Datagram Protocol using SMIv2 K. McCloghrie, Ed.

RFC 2018
TCP Selective Acknowledgement Options M. Mathis, J. Mahdavi, S. Floyd, A. Romanow

RFC 2026
The Internet Standards Process — Revision 3 S. Bradner

RFC 2030
Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI D. Mills

RFC 2033
Local Mail Transfer Protocol J. Myers

RFC 2034
*SMTP Service Extension for Returning Enhanced Error Codes*N. Freed

RFC 2040
*The RC5, RC5–CBC, RC-5–CBC-Pad, and RC5–CTS Algorithms*R. Baldwin, R. Rivest

RFC 2045
Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies N. Freed, N. Borenstein

RFC 2052
A DNS RR for specifying the location of services (DNS SRV) A. Gulbrandsen, P. Vixie

RFC 2065
Domain Name System Security Extensions D. Eastlake 3rd, C. Kaufman

RFC 2066
TELNET CHARSET Option R. Gellens

RFC 2080
RIPng for IPv6 G. Malkin, R. Minnear

RFC 2096
IP Forwarding Table MIB F. Baker

RFC 2104
HMAC: Keyed-Hashing for Message Authentication H. Krawczyk, M. Bellare, R. Canetti

RFC 2119
Keywords for use in RFCs to Indicate Requirement Levels S. Bradner

RFC 2133
Basic Socket Interface Extensions for IPv6 R. Gilligan, S. Thomson, J. Bound, W. Stevens

RFC 2136
Dynamic Updates in the Domain Name System (DNS UPDATE) P. Vixie, Ed., S. Thomson, Y. Rekhter, J. Bound

RFC 2137
Secure Domain Name System Dynamic Update D. Eastlake 3rd

RFC 2163
Using the Internet DNS to Distribute MIXER Conformant Global Address Mapping (MCGAM) C. Allocchio

RFC 2168
Resolution of Uniform Resource Identifiers using the Domain Name System R. Daniel, M. Mealling

RFC 2178
OSPF Version 2 J. Moy

RFC 2181
Clarifications to the DNS Specification R. Elz, R. Bush

RFC 2205
Resource ReSerVation Protocol (RSVP)—Version 1 Functional Specification R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, S. Jamin

RFC 2210
The Use of RSVP with IETF Integrated Services J. Wroclawski

RFC 2211
Specification of the Controlled-Load Network Element Service J. Wroclawski

RFC 2212
Specification of Guaranteed Quality of Service S. Shenker, C. Partridge, R. Guerin

RFC 2215
General Characterization Parameters for Integrated Service Network Elements S. Shenker, J. Wroclawski

RFC 2217
Telnet Com Port Control Option G. Clarke

RFC 2219
Use of DNS Aliases for Network Services M. Hamilton, R. Wright

RFC 2228
FTP Security Extensions M. Horowitz, S. Lunt

RFC 2230
Key Exchange Delegation Record for the DNS R. Atkinson

RFC 2233
The Interfaces Group MIB using SMIv2 K. McCloghrie, F. Kastenholz

RFC 2240
A Legal Basis for Domain Name Allocation O. Vaughn

RFC 2246
The TLS Protocol Version 1.0 T. Dierks, C. Allen

RFC 2251
Lightweight Directory Access Protocol (v3) M. Wahl, T. Howes, S. Kille

RFC 2253
Lightweight Directory Access Protocol (v3): UTF-8 String Representation of Distinguished Names M. Wahl, S. Kille, T. Howes

RFC 2254
The String Representation of LDAP Search Filters T. Howes

RFC 2261
An Architecture for Describing SNMP Management Frameworks D. Harrington, R. Presuhn, B. Wijnen

RFC 2262
Message Processing and Dispatching for the Simple Network Management Protocol (SNMP) J. Case, D. Harrington, R. Presuhn, B. Wijnen

RFC 2271
An Architecture for Describing SNMP Management Frameworks D. Harrington, R. Presuhn, B. Wijnen

RFC 2273
SNMPv3 Applications D. Levi, P. Meyer, B. Stewartz

RFC 2274
User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3) U. Blumenthal, B. Wijnen

RFC 2275
View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP) B. Wijnen, R. Presuhn, K. McCloghrie

RFC 2279
UTF-8, a transformation format of ISO 10646 F. Yergeau

RFC 2292
Advanced Sockets API for IPv6 W. Stevens, M. Thomas

RFC 2308
Negative Caching of DNS Queries (DNS NCACHE) M. Andrews

RFC 2317
Classless IN-ADDR.ARPA delegation H. Eidnes, G. de Groot, P. Vixie

RFC 2320
Definitions of Managed Objects for Classical IP and ARP Over ATM Using SMIv2 (IPOA-MIB) M. Greene, J. Luciani, K. White, T. Kuo

RFC 2328
OSPF Version 2 J. Moy

RFC 2345
Domain Names and Company Name Retrieval J. Klensin, T. Wolf, G. Oglesby

RFC 2352
A Convention for Using Legal Names as Domain Names O. Vaughn

RFC 2355
TN3270 Enhancements B. Kelly

RFC 2358
Definitions of Managed Objects for the Ethernet-like Interface Types J. Flick, J. Johnson

RFC 2373
IP Version 6 Addressing Architecture R. Hinden, S. Deering

RFC 2374
An IPv6 Aggregatable Global Unicast Address Format R. Hinden, M. O'Dell, S. Deering

RFC 2375
IPv6 Multicast Address Assignments R. Hinden, S. Deering

RFC 2385
Protection of BGP Sessions via the TCP MD5 Signature Option A. Hefferman

RFC 2389
Feature negotiation mechanism for the File Transfer Protocol P. Hethmon, R. Elz

RFC 2401
Security Architecture for Internet Protocol S. Kent, R. Atkinson

RFC 2402
IP Authentication Header S. Kent, R. Atkinson

RFC 2403
The Use of HMAC-MD5–96 within ESP and AH C. Madson, R. Glenn

RFC 2404
The Use of HMAC-SHA–1–96 within ESP and AH C. Madson, R. Glenn

RFC 2405
The ESP DES-CBC Cipher Algorithm With Explicit IV C. Madson, N. Doraswamy

RFC 2406
IP Encapsulating Security Payload (ESP) S. Kent, R. Atkinson

RFC 2407
*The Internet IP Security Domain of Interpretation for ISAKMP*D. Piper

RFC 2408
Internet Security Association and Key Management Protocol (ISAKMP) D. Maughan, M. Schertler, M. Schneider, J. Turner

RFC 2409
The Internet Key Exchange (IKE) D. Harkins, D. Carrel

RFC 2410
The NULL Encryption Algorithm and Its Use With IPsec R. Glenn, S. Kent,

RFC 2428
FTP Extensions for IPv6 and NATs M. Allman, S. Ostermann, C. Metz

RFC 2445
Internet Calendaring and Scheduling Core Object Specification (iCalendar) F. Dawson, D. Stenerson

RFC 2459
Internet X.509 Public Key Infrastructure Certificate and CRL Profile R. Housley, W. Ford, W. Polk, D. Solo

RFC 2460
Internet Protocol, Version 6 (IPv6) Specification S. Deering, R. Hinden

RFC 2461
Neighbor Discovery for IP Version 6 (IPv6) T. Narten, E. Nordmark, W. Simpson

RFC 2462
IPv6 Stateless Address Autoconfiguration S. Thomson, T. Narten

RFC 2463
Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification A. Conta, S. Deering

RFC 2464
Transmission of IPv6 Packets over Ethernet Networks M. Crawford

RFC 2466
Management Information Base for IP Version 6: ICMPv6 Group D. Haskin, S. Onishi

RFC 2476
Message Submission R. Gellens, J. Klensin

RFC 2487
SMTP Service Extension for Secure SMTP over TLS P. Hoffman

RFC 2505
Anti-Spam Recommendations for SMTP MTAs G. Lindberg

RFC 2523
Photuris: Extended Schemes and Attributes P. Karn, W. Simpson

RFC 2535
Domain Name System Security Extensions D. Eastlake 3rd

RFC 2538
Storing Certificates in the Domain Name System (DNS) D. Eastlake 3rd, O. Gudmundsson

RFC 2539
Storage of Diffie-Hellman Keys in the Domain Name System (DNS) D. Eastlake 3rd

RFC 2540
Detached Domain Name System (DNS) Information D. Eastlake 3rd

RFC 2554
SMTP Service Extension for Authentication J. Myers

RFC 2570
Introduction to Version 3 of the Internet-standard Network Management Framework J. Case, R. Mundy, D. Partain, B. Stewart

RFC 2571
An Architecture for Describing SNMP Management Frameworks B. Wijnen, D. Harrington, R. Presuhn

RFC 2572
Message Processing and Dispatching for the Simple Network Management Protocol (SNMP) J. Case, D. Harrington, R. Presuhn, B. Wijnen

RFC 2573
SNMP Applications D. Levi, P. Meyer, B. Stewart

RFC 2574
User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3) U. Blumenthal, B. Wijnen

RFC 2575
View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP) B. Wijnen, R. Presuhn, K. McCloghrie

RFC 2576
Co-Existence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework R. Frye, D. Levi, S. Routhier, B. Wijnen

RFC 2578
Structure of Management Information Version 2 (SMIv2) K. McCloghrie, D. Perkins, J. Schoenwaelder

RFC 2579
Textual Conventions for SMIv2 K. McCloghrie, D. Perkins, J. Schoenwaelder

RFC 2580
Conformance Statements for SMIv2 K. McCloghrie, D. Perkins, J. Schoenwaelder

RFC 2581
TCP Congestion Control M. Allman, V. Paxson, W. Stevens

RFC 2583
Guidelines for Next Hop Client (NHC) Developers R. Carlson, L. Winkler

RFC 2591
Definitions of Managed Objects for Scheduling Management Operations D. Levi, J. Schoenwaelder

RFC 2625
IP and ARP over Fibre Channel M. Rajagopal, R. Bhagwat, W. Rickard

RFC 2635
Don't SPEW A Set of Guidelines for Mass Unsolicited Mailings and Postings (spam)* S. Hambridge, A. Lunde

RFC 2637
Point-to-Point Tunneling Protocol K. Hamzeh, G. Pall, W. Verthein, J. Taarud, W. Little, G. Zorn

RFC 2640
Internationalization of the File Transfer Protocol B. Curtin

RFC 2665
Definitions of Managed Objects for the Ethernet-like Interface Types J. Flick, J. Johnson

RFC 2671
Extension Mechanisms for DNS (EDNS0) P. Vixie

RFC 2672
Non-Terminal DNS Name Redirection M. Crawford

RFC 2675
IPv6 Jumbograms D. Borman, S. Deering, R. Hinden

RFC 2710
Multicast Listener Discovery (MLD) for IPv6 S. Deering, W. Fenner, B. Haberman

RFC 2711
IPv6 Router Alert Option C. Partridge, A. Jackson

RFC 2740
OSPF for IPv6 R. Coltun, D. Ferguson, J. Moy

RFC 2753
A Framework for Policy-based Admission Control R. Yavatkar, D. Pendarakis, R. Guerin

RFC 2782
A DNS RR for specifying the location of services (DNS SRV) A. Gubrandsen, P. Vixix, L. Esibov

RFC 2821
Simple Mail Transfer Protocol J. Klensin, Ed.

RFC 2822
Internet Message Format P. Resnick, Ed.

RFC 2840
TELNET KERMIT OPTION J. Altman, F. da Cruz

RFC 2845
Secret Key Transaction Authentication for DNS (TSIG) P. Vixie, O. Gudmundsson, D. Eastlake 3rd, B. Wellington

RFC 2851
Textual Conventions for Internet Network Addresses M. Daniele, B. Haberman, S. Routhier, J. Schoenwaelder

RFC 2852
Deliver By SMTP Service Extension D. Newman

RFC 2874
DNS Extensions to Support IPv6 Address Aggregation and Renumbering M. Crawford, C. Huitema

RFC 2915
The Naming Authority Pointer (NAPTR) DNS Resource Record M. Mealling, R. Daniel

RFC 2920
SMTP Service Extension for Command Pipelining N. Freed

RFC 2930
Secret Key Establishment for DNS (TKEY RR) D. Eastlake, 3rd

RFC 2941
Telnet Authentication Option T. Ts'o, ed., J. Altman

RFC 2942
Telnet Authentication: Kerberos Version 5 T. Ts'o

RFC 2946
Telnet Data Encryption Option T. Ts'o

RFC 2952
Telnet Encryption: DES 64 bit Cipher Feedback T. Ts'o

RFC 2953
Telnet Encryption: DES 64 bit Output Feedback T. Ts'o

RFC 2992
Analysis of an Equal-Cost Multi-Path Algorithm C. Hopps

RFC 3019
IP Version 6 Management Information Base for The Multicast Listener Discovery Protocol B. Haberman, R. Worzella

RFC 3060
Policy Core Information Model—Version 1 Specification B. Moore, E. Ellesson, J. Strassner, A. Westerinen

RFC 3152
Delegation of IP6.ARPA R. Bush

RFC 3164
The BSD Syslog Protocol C. Lonvick

RFC 3207
SMTP Service Extension for Secure SMTP over Transport Layer Security P. Hoffman

RFC 3226
DNSSEC and IPv6 A6 aware server/resolver message size requirements O. Gudmundsson

RFC 3291
Textual Conventions for Internet Network Addresses M. Daniele, B. Haberman, S. Routhier, J. Schoenwaelder

RFC 3363
Representing Internet Protocol version 6 (IPv6) Addresses in the Domain Name System R. Bush, A. Durand, B. Fink, O. Gudmundsson, T. Hain

RFC 3376
Internet Group Management Protocol, Version 3 B. Cain, S. Deering, I. Kouvelas, B. Fenner, A. Thyagarajan

RFC 3390
Increasing TCP's Initial Window M. Allman, S. Floyd, C. Partridge

RFC 3410
Introduction and Applicability Statements for Internet-Standard Management Framework J. Case, R. Mundy, D. Partain, B. Stewart

RFC 3411
An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks D. Harrington, R. Presuhn, B. Wijnen

RFC 3412
Message Processing and Dispatching for the Simple Network Management Protocol (SNMP) J. Case, D. Harrington, R. Presuhn, B. Wijnen

RFC 3413
Simple Network Management Protocol (SNMP) Applications D. Levi, P. Meyer, B. Stewart

RFC 3414
User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3) U. Blumenthal, B. Wijnen

RFC 3415
View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP) B. Wijnen, R. Presuhn, K. McCloghrie

RFC 3416
Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP) R. Presuhn, J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 3417
Transport Mappings for the Simple Network Management Protocol (SNMP) R. Presuhn, J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 3418
Management Information Base (MIB) for the Simple Network Management Protocol (SNMP) R. Presuhn, J. Case, K. McCloghrie, M. Rose, S. Waldbusser

RFC 3419
Textual Conventions for Transport Addresses M. Daniele, J. Schoenwaelder

RFC 3484
Default Address Selection for Internet Protocol version 6 (IPv6) R. Draves

RFC 3493
Basic Socket Interface Extensions for IPv6 R. Gilligan, S. Thomson, J. Bound, J. McCann, W. Stevens

RFC 3513
Internet Protocol Version 6 (IPv6) Addressing Architecture R. Hinden, S. Deering

RFC 3526
More Modular Exponential (MODP) Diffie-Hellman groups for Internet Key Exchange (IKE) T. Kivinen, M. Kojo

RFC 3542
Advanced Sockets Application Programming Interface (API) for IPv6 W. Richard Stevens, M. Thomas, E. Nordmark, T. Jinmei

RFC 3566
The AES-XCBC-MAC-96 Algorithm and Its Use With IPsec S. Frankel, H. Herbert

RFC 3569
An Overview of Source-Specific Multicast (SSM) S. Bhattacharyya, Ed.

RFC 3584
Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework R. Frye, D. Levi, S. Routhier, B. Wijnen

RFC 3602
The AES-CBC Cipher Algorithm and Its Use with IPsec S. Frankel, R. Glenn, S. Kelly

RFC 3629
UTF-8, a transformation format of ISO 10646 R. Kermode, C. Vicisano

RFC 3658
Delegation Signer (DS) Resource Record (RR) O. Gudmundsson

RFC 3678
Socket Interface Extensions for Multicast Source Filters D. Thaler, B. Fenner, B. Quinn

RFC 3715
IPsec-Network Address Translation (NAT) Compatibility Requirements B. Aboba, W. Dixon

RFC 3810
Multicast Listener Discovery Version 2 (MLDv2) for IPv6 R. Vida, Ed., L. Costa, Ed.

RFC 3826
The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model U. Blumenthal, F. Maino, K McCloghrie.

RFC 3947
Negotiation of NAT-Traversal in the IKE T. Kivinen, B. Swander, A. Huttunen, V. Volpe

RFC 3948
UDP Encapsulation of IPsec ESP Packets A. Huttunen, B. Swander, V. Volpe, L. DiBurro, M. Stenberg

RFC 4001
Textual Conventions for Internet Network Addresses M. Daniele, B. Haberman, S. Routhier, J. Schoenwaelder

RFC 4007
IPv6 Scoped Address Architecture S. Deering, B. Haberman, T. Jinmei, E. Nordmark, B. Zill

RFC 4022
Management Information Base for the Transmission Control Protocol (TCP) R. Raghunarayan

RFC 4106
The Use of Galois/Counter Mode (GCM) in IPsec Encapsulating Security Payload (ESP) J. Viega, D. McGrew

RFC 4109
Algorithms for Internet Key Exchange version 1 (IKEv1) P. Hoffman

RFC 4113
Management Information Base for the User Datagram Protocol (UDP) B. Fenner, J. Flick

RFC 4191
Default Router Preferences and More-Specific Routes R. Draves, D. Thaler

RFC 4217
Securing FTP with TLS P. Ford-Hutchinson

RFC 4292
IP Forwarding Table MIB B. Haberman

RFC 4293
Management Information Base for the Internet Protocol (IP) S. Routhier

RFC 4301
Security Architecture for the Internet Protocol S. Kent, K. Seo

RFC 4302
IP Authentication Header S. Kent

RFC 4303
IP Encapsulating Security Payload (ESP) S. Kent

RFC 4304
Extended Sequence Number (ESN) Addendum to IPsec Domain of Interpretation (DOI) for Internet Security Association and Key Management Protocol (ISAKMP) S. Kent

RFC 4307
Cryptographic Algorithms for Use in the Internet Key Exchange Version 2 (IKEv2) J. Schiller

RFC 4308
Cryptographic Suites for IPsec P. Hoffman

RFC 4434
The AES-XCBC-PRF-128 Algorithm for the Internet Key Exchange Protocol P. Hoffman

RFC 4443
Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification A. Conta, S. Deering

RFC 4552
Authentication/Confidentiality for OSPFv3 M. Gupta, N. Melam

RFC 4678
Server/Application State Protocol v1 A. Bivens

RFC 4753
ECP Groups for IKE and IKEv2 D. Fu, J. Solinas

RFC 4754
IKE and IKEv2 Authentication Using the Elliptic Curve Digital Signature Algorithm (ECDSA) D. Fu, J. Solinas

RFC 4809
Requirements for an IPsec Certificate Management Profile C. Bonatti, Ed., S. Turner, Ed., G. Lebovitz, Ed.

RFC 4835
Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH) V. Manral

RFC 4862
IPv6 Stateless Address Autoconfiguration S. Thomson, T. Narten, T. Jinmei

RFC 4868
Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec S. Kelly, S. Frankel

RFC 4869
Suite B Cryptographic Suites for IPsec L. Law, J. Solinas

RFC 4941
Privacy Extensions for Stateless Address Autoconfiguration in IPv6 T. Narten, R. Draves, S. Krishnan

RFC 4945
The Internet IP Security PKI Profile of IKEv1/ISAKMP, IKEv2, and PKIX B. Korver

RFC 5014
IPv6 Socket API for Source Address Selection E. Nordmark, S. Chakrabarti, J. Laganier

RFC 5095
Deprecation of Type 0 Routing Headers in IPv6 J. Abley, P. Savola, G. Neville-Neil

RFC 5175
IPv6 Router Advertisement Flags Option B. Haberman, Ed., R. Hinden

RFC 5282
Using Authenticated Encryption Algorithms with the Encrypted Payload of the Internet Key Exchange version 2 (IKEv2) Protocol D. Black, D. McGrew

RFC 5996
Internet Key Exchange Protocol Version 2 (IKEv2) C. Kaufman, P. Hoffman, Y. Nir, P. Eronen

RFC 7627
Transport Layer Security (TLS) Session Hash and Extended Master Secret Extension K. Bhargavan, A. Delignat-Lavaud, A. Pironti, Inria Paris-Rocquencourt, A. Langley, M. Ray

RFC 8446
The Transport Layer Security (TLS) Protocol Version 1.3 E. Rescorla

Internet drafts

Internet drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Other groups can also distribute working documents as Internet drafts. You can see Internet drafts at http://www.ietf.org/ID.html.

Appendix H. Accessibility

Publications for this product are offered in Adobe Portable Document Format (PDF) and should be compliant with accessibility standards. If you experience difficulties when using PDF files, you can view the information through the z/OS Internet Library website http://www.ibm.com/systems/z/os/zos/library/bkserv/ or IBM Documentation https://www.ibm.com/docs/en. If you continue to experience problems, send a message to Contact z/OS web page(www.ibm.com/systems/z/os/zos/webqs.html) or write to:

IBM Corporation
Attention: MHVRCFS Reader Comments
Department H6MA, Building 707
2455 South Road
Poughkeepsie, NY 12601-5400
USA

Accessibility features help a user who has a physical disability, such as restricted mobility or limited vision, to use software products successfully. The major accessibility features in z/OS enable users to:

- Use assistive technologies such as screen readers and screen magnifier software
- Operate specific or equivalent features using only the keyboard
- Customize display attributes such as color, contrast, and font size

Using assistive technologies

Assistive technology products, such as screen readers, function with the user interfaces found in z/OS. Consult the assistive technology documentation for specific information when using such products to access z/OS interfaces.

Keyboard navigation of the user interface

Users can access z/OS user interfaces using TSO/E or ISPF. See z/OS TSO/E Primer, z/OS TSO/E User's Guide, and z/OS ISPF User's Guide Vol I for information about accessing TSO/E and ISPF interfaces. These guides describe how to use TSO/E and ISPF, including the use of keyboard shortcuts or function keys (PF keys). Each guide includes the default settings for the PF keys and explains how to modify their functions.

z/OS information

One exception is command syntax that is published in railroad track format, which is accessible using screen readers with IBM Documentation, as described in “Dotted decimal syntax diagrams” on page 1429.

Dotted decimal syntax diagrams

Syntax diagrams are provided in dotted decimal format for users accessing IBM Documentation using a screen reader. In dotted decimal format, each syntax element is written on a separate line. If two or more syntax elements are always present together (or always absent together), they can appear on the same line, because they can be considered as a single compound syntax element.

Each line starts with a dotted decimal number; for example, 3 or 3.1 or 3.1.1. To hear these numbers correctly, make sure that your screen reader is set to read out punctuation. All the syntax elements that have the same dotted decimal number (for example, all the syntax elements that have the number 3.1) are mutually exclusive alternatives. If you hear the lines 3.1 USERID and 3.1 SYSTEMID, you know that your syntax can include either USERID or SYSTEMID, but not both.

© Copyright IBM Corp. 2000, 2022

The dotted decimal numbering level denotes the level of nesting. For example, if a syntax element with dotted decimal number 3 is followed by a series of syntax elements with dotted decimal number 3.1, all the syntax elements numbered 3.1 are subordinate to the syntax element numbered 3.

Certain words and symbols are used next to the dotted decimal numbers to add information about the syntax elements. Occasionally, these words and symbols might occur at the beginning of the element itself. For ease of identification, if the word or symbol is a part of the syntax element, it is preceded by the backslash (\) character. The * symbol can be used next to a dotted decimal number to indicate that the syntax element repeats. For example, syntax element *FILE with dotted decimal number 3 is given the format 3 * FILE. Format 3* FILE indicates that syntax element FILE repeats. Format 3* * FILE indicates that syntax element * FILE repeats.

Characters such as commas, which are used to separate a string of syntax elements, are shown in the syntax just before the items they separate. These characters can appear on the same line as each item, or on a separate line with the same dotted decimal number as the relevant items. The line can also show another symbol giving information about the syntax elements. For example, the lines 5.1*, 5.1 LASTRUN, and 5.1 DELETE mean that if you use more than one of the LASTRUN and DELETE syntax elements, the elements must be separated by a comma. If no separator is given, assume that you use a blank to separate each syntax element.

If a syntax element is preceded by the % symbol, this indicates a reference that is defined elsewhere. The string following the % symbol is the name of a syntax fragment rather than a literal. For example, the line 2.1 %OP1 means that you should see separate syntax fragment OP1.

The following words and symbols are used next to the dotted decimal numbers:

- A question mark (?) means an optional syntax element. A dotted decimal number followed by the ? symbol indicates that all the syntax elements with a corresponding dotted decimal number, and any subordinate syntax elements, are optional. If there is only one syntax element with a dotted decimal number, the ? symbol is displayed on the same line as the syntax element, (for example 5? NOTIFY). If there is more than one syntax element with a dotted decimal number, the ? symbol is displayed on a line by itself, followed by the syntax elements that are optional. For example, if you hear the lines 5 ?, 5 NOTIFY, and 5 UPDATE, you know that syntax elements NOTIFY and UPDATE are optional; that is, you can choose one or none of them. The ? symbol is equivalent to a bypass line in a railroad diagram.
- An exclamation mark (!) means a default syntax element. A dotted decimal number followed by the ! symbol and a syntax element indicate that the syntax element is the default option for all syntax elements that share the same dotted decimal number. Only one of the syntax elements that share the same dotted decimal number can specify a ! symbol. For example, if you hear the lines 2? FILE, 2.1! (KEEP), and 2.1 (DELETE), you know that (KEEP) is the default option for the FILE keyword. In this example, if you include the FILE keyword but do not specify an option, default option KEEP will be applied. A default option also applies to the next higher dotted decimal number. In this example, if the FILE keyword is omitted, default FILE(KEEP) is used. However, if you hear the lines 2? FILE, 2.1, 2.1.1! (KEEP), and 2.1.1 (DELETE), the default option KEEP applies only to the next higher dotted decimal number, 2.1 (which does not have an associated keyword), and does not apply to 2? FILE. Nothing is used if the keyword FILE is omitted.
- An asterisk (*) means a syntax element that can be repeated 0 or more times. A dotted decimal number followed by the * symbol indicates that this syntax element can be used zero or more times; that is, it is optional and can be repeated. For example, if you hear the line 5.1* data area, you know that you can include one data area, more than one data area, or no data area. If you hear the lines 3*, 3 HOST, and 3 STATE, you know that you can include HOST, STATE, both together, or nothing.

  Notes:

  1. If a dotted decimal number has an asterisk (*) next to it and there is only one item with that dotted decimal number, you can repeat that same item more than once.
  2. If a dotted decimal number has an asterisk next to it and several items have that dotted decimal number, you can use more than one item from the list, but you cannot use the items more than once each. In the previous example, you could write HOST STATE, but you could not write HOST HOST.
  3. The * symbol is equivalent to a loop-back line in a railroad syntax diagram.

- + means a syntax element that must be included one or more times. A dotted decimal number followed by the + symbol indicates that this syntax element must be included one or more times; that is, it must be included at least once and can be repeated. For example, if you hear the line 6.1+ data area, you must include at least one data area. If you hear the lines 2+, 2 HOST, and 2 STATE, you know that you must include HOST, STATE, or both. Similar to the * symbol, the + symbol can only repeat a particular item if it is the only item with that dotted decimal number. The + symbol, like the * symbol, is equivalent to a loop-back line in a railroad syntax diagram.

Notices

This information was developed for products and services that are offered in the USA or elsewhere.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing IBM Corporation North Castle Drive, MD-NC119 Armonk, NY 10504-1785 United States of America

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing Legal and Intellectual Property Law IBM Japan Ltd. 19-21, Nihonbashi-Hakozakicho, Chuo-ku Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

This information could include missing, incorrect, or broken hyperlinks. Hyperlinks are maintained in only the HTML plug-in output for the IBM Documentation. Use of hyperlinks in other output formats of this information is at your own risk.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation Site Counsel 2455 South Road Poughkeepsie, NY 12601-5400 USA

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

© Copyright IBM Corp. 2000, 2022

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

IBM Online Privacy Statement

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user, or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session cookies that collect each user's name, email address, phone number, or other personally identifiable information for purposes of enhanced user usability and single sign-on configuration. These cookies can be disabled, but disabling them will also eliminate the functionality they enable.

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at ibm.com®/privacy and IBM's Online Privacy Statement at ibm.com/privacy/details in the section entitled "Cookies, Web Beacons and Other Technologies," and the "IBM Software Products and Software-as-a-Service Privacy Statement" at ibm.com/software/info/product-privacy.

Policy for unsupported hardware

Various z/OS elements, such as DFSMS, JES2, JES3, and MVS, contain code that supports specific hardware servers or devices. In some cases, this device-related element support remains in the product even after the hardware devices pass their announced End of Service date. z/OS may continue to service element code; however, it will not provide service related to unsupported hardware devices. Software problems related to these devices will not be accepted for service, and current service activity will cease if a problem is determined to be associated with out-of-support devices. In such cases, fixes will not be issued.

Minimum supported hardware

The minimum supported hardware for z/OS releases identified in z/OS announcements can subsequently change when service for particular servers or devices is withdrawn. Likewise, the levels of other software products supported on a particular release of z/OS are subject to the service support lifecycle of those products. Therefore, z/OS and its product publications (for example, panels, samples, messages, and product documentation) can include references to hardware and software that is no longer supported.

- For information about software support lifecycle, see: IBM Lifecycle Support for z/OS (www.ibm.com/software/support/systemsz/lifecycle)
- For information about currently-supported IBM hardware, contact your IBM representative.

Policy for unsupported hardware

Various z/OS elements, such as DFSMS, HCD, JES2, JES3, and MVS, contain code that supports specific hardware servers or devices. In some cases, this device-related element support remains in the product even after the hardware devices pass their announced End of Service date. z/OS may continue to service element code; however, it will not provide service related to unsupported hardware devices. Software problems related to these devices will not be accepted for service, and current service activity will cease if a problem is determined to be associated with out-of-support devices. In such cases, fixes will not be issued.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at Copyright and trademark information at www.ibm.com/legal/copytrade.shtml.

Bibliography

This bibliography contains descriptions of the documents in the z/OS Communications Server library.

z/OS Communications Server documentation is available online at the z/OS Internet Library web page at http://www.ibm.com/systems/z/os/zos/library/bkserv/.

z/OS Communications Server library updates

Updates to documents are also available on RETAIN and in information APARs (info APARs). Go to https://www.ibm.com/mysupport to view information APARs.

- z/OS Communications Server V2R1 New Function APAR Summary
- z/OS Communications Server V2R2 New Function APAR Summary
- z/OS Communications Server V2R3 New Function APAR Summary
- z/OS Communications Server V2R4 New Function APAR Summary

z/OS Communications Server information

z/OS Communications Server product information is grouped by task in the following tables.

Planning

Title	Number	Description
z/OS Communications Server: New Function Summary	GC27-3664	This document is intended to help you plan for new IP or SNA functions, whether you are migrating from a previous version or installing z/OS for the first time. It summarizes what is new in the release and identifies the suggested and required modifications needed to use the enhanced functions.
z/OS Communications Server: IPv6 Network and Application Design Guide	SC27-3663	This document is a high-level introduction to IPv6. It describes concepts of z/OS Communications Server's support of IPv6, coexistence with IPv4, and migration issues.

Resource definition, configuration, and tuning

Title	Number	Description
z/OS Communications Server: IP Configuration Guide	SC27-3650	This document describes the major concepts involved in understanding and configuring an IP network. Familiarity with the z/OS operating system, IP protocols, z/OS UNIX System Services, and IBM Time Sharing Option (TSO) is recommended. Use this document with the z/OS Communications Server: IP Configuration Reference.

© Copyright IBM Corp. 2000, 2022

Title	Number	Description
z/OS Communications Server: IP Configuration Reference	SC27-3651	This document presents information for people who want to administer and maintain IP. Use this document with the z/OS Communications Server: IP Configuration Guide. The information in this document includes: • TCP/IP configuration data sets • Configuration statements • Translation tables • Protocol number and port assignments
z/OS Communications Server: SNA Network Implementation Guide	SC27-3672	This document presents the major concepts involved in implementing an SNA network. Use this document with the z/OS Communications Server: SNA Resource Definition Reference.
z/OS Communications Server: SNA Resource Definition Reference	SC27-3675	This document describes each SNA definition statement, start option, and macroinstruction for user tables. It also describes NCP definition statements that affect SNA. Use this document with the z/OS Communications Server: SNA Network Implementation Guide.
z/OS Communications Server: SNA Resource Definition Samples	SC27-3676	This document contains sample definitions to help you implement SNA functions in your networks, and includes sample major node definitions.
z/OS Communications Server: IP Network Print Facility	SC27-3658	This document is for systems programmers and network administrators who need to prepare their network to route SNA, JES2, or JES3 printer output to remote printers using TCP/IP Services.

Operation

Title	Number	Description
z/OS Communications Server: IP User's Guide and Commands	SC27-3662	This document describes how to use TCP/IP applications. It contains requests with which a user can log on to a remote host using Telnet, transfer data sets using FTP, send electronic mail, print on remote printers, and authenticate network users.
z/OS Communications Server: IP System Administrator's Commands	SC27-3661	This document describes the functions and commands helpful in configuring or monitoring your system. It contains system administrator's commands, such as TSO NETSTAT, PING, TRACERTE and their UNIX counterparts. It also includes TSO and MVS commands commonly used during the IP configuration process.
z/OS Communications Server: SNA Operation	SC27-3673	This document serves as a reference for programmers and operators requiring detailed information about specific operator commands.
z/OS Communications Server: Quick Reference	SC27-3665	This document contains essential information about SNA and IP commands.

Customization

Title	Number	Description
z/OS Communications Server: SNA Customization	SC27-3666	This document enables you to customize SNA, and includes the following information: • Communication network management (CNM) routing table • Logon-interpret routine requirements • Logon manager installation-wide exit routine for the CLU search exit • TSO/SNA installation-wide exit routines • SNA installation-wide exit routines

Writing application programs

Title	Number	Description
z/OS Communications Server: IP Sockets Application Programming Interface Guide and Reference	SC27-3660	This document describes the syntax and semantics of program source code necessary to write your own application programming interface (API) into TCP/IP. You can use this interface as the communication base for writing your own client or server application. You can also use this document to adapt your existing applications to communicate with each other using sockets over TCP/IP.
z/OS Communications Server: IP CICS Sockets Guide	SC27-3649	This document is for programmers who want to set up, write application programs for, and diagnose problems with the socket interface for CICS using z/OS TCP/IP.
z/OS Communications Server: IP IMS Sockets Guide	SC27-3653	This document is for programmers who want application programs that use the IMS TCP/IP application development services provided by the TCP/IP Services of IBM.
z/OS Communications Server: IP Programmer's Guide and Reference	SC27-3659	This document describes the syntax and semantics of a set of high-level application functions that you can use to program your own applications in a TCP/IP environment. These functions provide support for application facilities, such as user authentication, distributed databases, distributed processing, network management, and device sharing. Familiarity with the z/OS operating system, TCP/IP protocols, and IBM Time Sharing Option (TSO) is recommended.
z/OS Communications Server: SNA Programming	SC27-3674	This document describes how to use SNA macroinstructions to send data to and receive data from (1) a terminal in either the same or a different domain, or (2) another application program in either the same or a different domain.
z/OS Communications Server: SNA Programmer's LU 6.2 Guide	SC27-3669	This document describes how to use the SNA LU 6.2 application programming interface for host application programs. This document applies to programs that use only LU 6.2 sessions or that use LU 6.2 sessions along with other session types. (Only LU 6.2 sessions are covered in this document.)
z/OS Communications Server: SNA Programmer's LU 6.2 Reference	SC27-3670	This document provides reference material for the SNA LU 6.2 programming interface for host application programs.

Title	Number	Description
z/OS Communications Server: CSM Guide	SC27-3647	This document describes how applications use the communications storage manager.

Diagnosis

Title	Number	Description
z/OS Communications Server: IP Diagnosis Guide	GC27-3652	This document explains how to diagnose TCP/IP problems and how to determine whether a specific problem is in the TCP/IP product code. It explains how to gather information for and describe problems to the IBM Software Support Center.
z/OS Communications Server: ACF/TAP Trace Analysis Handbook	GC27-3645	This document explains how to gather the trace data that is collected and stored in the host processor. It also explains how to use the Advanced Communications Function/Trace Analysis Program (ACF/TAP) service aid to produce reports for analyzing the trace data information.
z/OS Communications Server: SNA Diagnosis Vol 1, Techniques and Procedures and z/OS Communications Server: SNA Diagnosis Vol 2, FFST Dumps and the VIT	GC27-3667 GC27-3668	These documents help you identify an SNA problem, classify it, and collect information about it before you call the IBM Support Center. The information collected includes traces, dumps, and other problem documentation.
z/OS Communications Server: SNA Data Areas Volume 1 and z/OS Communications Server: SNA Data Areas Volume 2	GC31-6852 GC31-6853	These documents describe SNA data areas and can be used to read an SNA dump. They are intended for IBM programming service representatives and customer personnel who are diagnosing problems with SNA.

Messages and codes

Title	Number	Description
z/OS Communications Server: SNA Messages	SC27-3671	This document describes the ELM, IKT, IST, IUT, IVT, and USS messages. Other information in this document includes: • Command and RU types in SNA messages • Node and ID types in SNA messages • Supplemental message-related information
z/OS Communications Server: IP Messages Volume 1 (EZA)	SC27-3654	This volume contains TCP/IP messages beginning with EZA.
z/OS Communications Server: IP Messages Volume 2 (EZB, EZD)	SC27-3655	This volume contains TCP/IP messages beginning with EZB or EZD.
z/OS Communications Server: IP Messages Volume 3 (EZY)	SC27-3656	This volume contains TCP/IP messages beginning with EZY.
z/OS Communications Server: IP Messages Volume 4 (EZZ, SNM)	SC27-3657	This volume contains TCP/IP messages beginning with EZZ and SNM.
z/OS Communications Server: IP and SNA Codes	SC27-3648	This document describes codes and other information that appear in z/OS Communications Server messages.

Index

Special Characters

/etc/ftp.data 676
/etc/hosts
  accessing HOSTS.SITEINFO 779
/etc/inetd.conf
  adding applications to 1327
  configuring z/OS UNIX REXECD 1313
  definition 1313
  setting traces in 1327
/etc/osnmpd.data 19
/etc/pagent.conf 825
/etc/protocol 781
/etc/pw.src 20
/etc/resolv.conf
  overview 234
  use of system names in 235
/etc/services
  defining ports for RSHD 1314
  defining ports for z/OS UNIX REXECD 1313
  specifying syslog service 232
/etc/snmpd.boots 1250
/etc/snmpd.conf 1250
/etc/snmptrap.dest 22
/etc/syslog.conf
  configuring for syslogd 213, 1328
  for FTP messages and traces 665
  overview 30
/etc/trapfwd.conf 1265

Numerics

328x printer support 594

A

access control
  /OS Encryption Readiness Technology (zERT) 147
  Fast Response Cache Accelerator 146
  IPSec network management interface (NMI) 148
  Netstat 145
  network 140
  port 137
  SMF information service 147
  stack 137
  TCP connection information service 146
  TCP/IP packet trace service 146
accessibility 1429
accounting, SMF records
  FTP 32, 685, 708
  PROFILE.TCPIP 240
  syslogd 228
  Telnet 32, 648
address selection, source IP 244
ADNR 1195
advertisements, router 357
AF_INET physical file system
  common 8
  integrated sockets 8
AF_INET problems 116
anchor filters 933
anonymous logins, configuring FTP for 714
APIs (application programming interfaces) 7
application programming interfaces, types in z/OS Communications Server, see also APIs 7
Application Transparent Transport Layer Security (AT-TLS) 1121
applications
  configuration files for TCP/IP 13, 25
  planning scenarios for multiple instances 373
applications, functions and protocols
  Character Generator protocol 1323
  Discard protocol 1323
  Echo protocol 1323
  Network Computing System (NCS) 1280
  Portmapper 1273, 1275
  Remote Execution Protocol Daemon (REXECD) 1309, 1313
  Remote Printing 1269
  Remote Procedure Call (RPC)
    Network Computing System (NCS) 1280
    Portmapper 1274
  Simple Network Management Protocol (SNMP) 1237
ARM (automatic restart manager) 29
ARPTO (IPCONFIG ARPTO) 240
AS (autonomous system)
  definition 279
assembler callable services, z/OS UNIX, general description 9
AT-TLS 1121
ATCCON member of VTAMLST 120
authorization, TCP/IP started task user ID 36
authorization, z/OS UNIX superuser 36
autoconfiguration, stateless 258
AUTOLOG 664
automated domain name registration 1195
automated takeover, VIPA 370
automatic fast path support 42
automatic restart manager (ARM) 29
AUTOMOUNT 676
autonomous system, see also AS 279
AUTORECALL 676

B

backing up an MVS host with VIPA 368
banner page 1271
BLKSIZE 676
BPX.DAEMON FACILITY class profile 37, 38
BPX.SMF 225
BPXPRMxx

BPXPRMxx *(continued)*
CINET configuration 54
BPXPRMxx, for defining z/OS UNIX environment 39
BPXPRMxx, role in AF_INET problems 116
bridge
sendmail 1297, 1298
BUFNO 676

C

C sockets 8
cataloged procedures
MISCSERV (MISCSERV) 1324
RXSERVE (RXPROC) 1309, 1313
SNMPD (SNMPDPRC) 1255
SNMPQE (SNMPPROC) 1258
channel-to-channel (CTC), general description 5
CICS (customer information control system) sockets 8
Cisco
Content Switching Module (CSM) 470
CLAWUSEDOUBLENOP 241
code page conversion
control connection 681
data connection 681
table priority, control connection 681
table priority, data connection 681
code page conversion, FTP 680
commands
MODIFY (MVS)
Remote Execution server 1311
START (MVS) 121
common AF_INET
access by APIs 8
general description 7
Communications Server for z/OS, online information xlvi
communications storage manager, general description 5
component trace, customizing 114
CONDDISP 676
configuration
data set security 150
configuration considerations
SMC 518
configuration data sets
ETCRPC 1274
HOSTS 265
SAMPPROF 237
configuring
data set naming conventions 13
dynamic VIPA 372
files for TCP/IP applications 25
files for the TCP/IP stack 23
resolver environment variables 775
searching for data sets 13
SNMP for z/OS UNIX 1237
SNTPD 1305
TIMED 1303
verifying for dynamic VIPAs 420
Configuring the z/OS UNIX Telnet server 655
considerations,
authentication 1314
Kerberos 1314
Content Switching Module (CSM), Cisco 470
control characters, TCP/IP messages 276
control connection, code page conversions 681
conversion characters, TCP/IP messages 276
conversion tables, control connection 681
cryptographic standards and FIPS 140 151
cryptography 151
CSM (communications storage manager), general description 5
CSM, Cisco 470
CSSMTP 1297, 1298
CTC (channel-to-channel), general description 5
CTRACE keyword 114
customer information control system sockets, general description, see also CICS 8
customizing TCP/IP messages 271

D

data connection
code page conversions 681
network transfer/file system conversion 681
data sets
dynamic allocation 13
naming conventions 13
overview 12
search order for 13
DATACLASS 676, 677
DATAGRAMFWD (IPCONFIG DATAGRAMFWD) 241, 418
Db2 708, 720
Db2 SQL
in FTP server 720
DB2PLAN 708
DCAS 1317
DCBDSN 677
DD cards 24
DEFAULTTCPIPDATA statement 741
Defense Manager daemon (DMD) 1107
defensive filtering 1107
Differentiated Services (DS)
Policies 831
Digital Certificate Access Server (DCAS) 1317
Direct Memory Access 503, 507
direct memory buffer 515
DIRECTORY 677
disability 1429
distributed DVIPA 361
DMB 515
DMD 1107
DNS (Domain Name System)
authoritative servers 731
caching-only servers 732
forwarders 732
master name servers 732
overview 729
secondary name servers 732
stealth server 733
DNS name server responsiveness
diagnosis 764
examples of 766
notifications for 763
DNS, online information xlvii
domain name registration, automated 1195
Domain Name System, see DNS 729
DPI (distributed prototcol interface) 1242
DSNLOAD 722
DSNTYPE 677

duplicate address detection 258
DVIPA takeover
  overview 396
  using IPSec with 398
DVIPSEC 396
dynamic filters 933
dynamic routes
  definition 279
dynamic routing
  IPv6 291
  using OMPROUTE 288
  versus static routing 281
dynamic VIPA
  1024 limit 369
  configuration 372, 418, 420
  MODDVIPA utility 376
  multiple application-instance scenario 372
  overview 361
  relationship to UDP 416
  resolving conflicts 399
  routing protocols 431
  unique application-instance scenario 372, 373
  use with OMPROUTE 295, 312
  verifying configuration using Netstat 423
  verifying in a sysplex 420
  within subnets 415
DYNAMICXCF 440
DYNAMICXCF (IPCONFIG DYNAMICXCF) 238, 241, 255, 385
DYNAMICXCF (IPCONFIG6 DYNAMICXCF) 256

E

EATTR 677
EGP (exterior gateway protocol)
  definition 279
Enterprise Extender
  overview 54
  VIPA considerations 364, 368
entry point name incorrect 116
environment variables
  for overriding default search order 13
  FTP server and 673
  OMPROUTE use of 300
  passing to syslogd process 224
  resolver configuration files and 775
  REXECD and 1314
ephemeral port selection 246
ETC.IPNODES 264, 267
ETC.SERVICES
  FTP and 665
Express Logon Feature (ELF)
  overview 1335
express logon services 1317
Extension Mechanisms for DNS standards (EDNS0) 770
exterior gateway protocol (EGP)
  definition 279
EZACFSM1 28
EZAFCCMD 711
EZAFCREP 712
EZAZSSI 117
EZBDVIPAvvtt 396, 399
EZBEPORTvvtt 390

F

fast path for socket applications 42
Fast Response Cache Accelerator access control 146
fault tolerance, interface layer for LANs 257
File systems, z/OS Communications Server TCP/IP 7
FTCHKCMD 709
FTCHKIP 709
FTCHKJES 710
FTCHKPWD 709
FTOEBIND 720
FTP
  /etc/syslog.conf 665
  accounting 32, 685
  anonymous 714, 718, 726
  APPEND 685
  AUTOLOG PORT KEEPALIVE 664
  cataloged procedure 665, 720
  CCXLATE 673
  code page conversion 680
  code page conversion for the control connection 681
  code page conversion for the data connection 681
  configuration statements, TCP/IP 664
  configuring with multiple stacks 675
  control connection, code page conversion 681
  control connection, conversion tables priority 681
  data connection, code page conversion 681
  data connection, conversion tables priority 681
  data translation 680
  Db2 720
  DELETE 685
  ENVAR 674
  environment variables for FTP server 673
  EZAFCCMD 711
  EZAFCREP 712
  FTCHKCMD 709
  FTCHKIP 709
  FTCHKJES 710
  FTCHKPWD 709
  FTP.DATA data set 676
  FTPOSTPR 710
  FTPSMFEX 708
  iconv function 680
  JES 713
  message catalogs, customizing 683
  priority for conversion tables, control connection 681
  priority for conversion tables, data connection 681
  RACF considerations 666
  RENAME 685
  RETRIEVE 685
  security considerations 666
  SMF configuration 685
  specifying attributes for new MVS data sets 677
  STORE 685
  STORE UNIQUE 685
  SURROGATE 714
  TCPIP.DATA 675
  TLS 686
  translation of data 680
  updating the FTP cataloged procedure 665
  user exit 708
  XLATE 673
FTP server
  preventing exploitation of 671

FTP.DATA
(FILETYPE=JES) 685
(FILETYPE=SEQ) 685
(FILETYPE=SQL) 685
ANONYMOUSHFSINFO 718
ANONYMOUSLOGINMSG 718
ANONYMOUSMVSINFO 718
ASATRANS 680
AUTOMOUNT 676
AUTORECALL 676
BANNER 718
BLKSIZE 676, 678
BLOCKSIZE 676
BUFNO 676
CONDDISP 676
CTRLCONN 680
data set attributes 676
DATACLASS 676–678
Db2 708
DB2PLAN 708
DBSUB 680
DCBDSN 677, 678
DIRECTORY 677–679
DSNTYPE 677–679
dynamic allocation 677
EATTR 677–679
ENCODING 680
EXTENSIONS UTF8 680
HSFINFO 718
JESGETBYDSN 708
JESINTERFACELEVEL 708
JESINTERFACELevel=2 713
JESLRECL 708
JESPUTGETTO 708
JESRECFM 708
LOGINMSG 718
LRECL 676–679
MBDATACONN 680
MBREQUIRELASTEOL 680
MBSENDEOL 680
MGMTCLASS 677, 678
MIGRATEVOL 677
MVSINFO 718
PDSTYPE 677–679
PORTCOMMAND 671, 672
PORTCOMMANDIPADDR 671, 672
PORTCOMMANDPORT 671, 672
PRIMARY 677–679
RECFM 677–679
RETPD 677–679
SBDATACONN 680
SBSENDEOL 680
SBSUB 680
SBSUBCHAR 680
search order 676
SECONDARY 677–679
SMFAPPE 685
SMFDEL 685
SMFEXIT 685
SMFJES 685
SMFLOGN 685
SMFREN 685
SMFRETR 685
SMFSQL 685
FTP.DATA *(continued)*
SMFSTOR 685
SMS 679
SPACETYPE 677, 678
SPREAD 708
SQLCOL 708
STORCLASS 677, 678
UCOUNT 677, 678
UCSHOSTCS 680
UCSSUB 680
UCSTRUNCT 680
UMASK 677
UNICODEFILESYSTEMBOM 680
UNITNAME 677, 678
VCOUNT 677, 678
VOLUME 677, 678
XLATE 685
FTPD 665
FTPOSTPR 710
FTPSMFEX 708
FTPSMFEX user exit 708

G

gateways
resolving names of 270
generic stack affinity 46
global TCPIP.DATA file 741
GLOBALTCPIPDATA statement 741

H

HCD, using 1339
hierarchical file system concepts 12
high availability, SMC-R 522
high-level qualifier (HLQ) 14
HiperSockets
concepts 75
HiperSockets Converged Interface 98, 99
Layer 2 88
Layer 3 88
Linux 95
performance 97
PNetID 98
Shared Memory Communications 98
virtual LAN 79
z/VM VSwitch bridge 95
HiperSockets Accelerator
efficient routing with 84
HLQ (high-level qualifier) 14
HOMETEST 270
HOSTALIASES 775
HOSTS.ADDRINFO
generating from HOSTS.LOCAL 265
HOSTS.LOCAL 264
HOSTS.SITEINFO
generating from HOSTS.LOCAL 265
verifying 270

I

I/O process model 5
IBM 10 GbE RoCE Express feature 6, 504

IBM Configuration Assistant for z/OS Communications Server
  AT-TLS 1122
  IDS 863
  IP security 892
  overview 789
  policy-based routing 350
  QoS 837
IBM Software Support Center, contacting xxxvii
IBM z/OS Management Facility
  AT-TLS 1122
  DMD 1116
  IDS 863
  overview 789
  policy-based routing 350
  QoS 837
ICMP (internet control message protocol), general description 6
iconv function 680
IDS
  defining policies 864
IGNOREREDIRECTS (IPCONFIG IGNOREREDIRECTS) 284
IGP (interior gateway protocol), definition 280
IKE daemon, preparing to run 1361
IMS sockets 8
in-addr.arpa domain, definition 729
InetD configuration file, setting up 1327
inetd listener program 37
Information APARs xliv
initialization failure 115
installing z/OS Communications Server 114
instances of TCPIP, considerations for multiple 45
interface takeover 257
interface-layer fault-tolerance for LANs 257
interior gateway protocol (IGP), definition 280
internet control message protocol (ICMP), general description 6
Internet protocol (IP), definition 4
Internet, finding z/OS information online xlvi
InterNetwork Information Center (InterNIC) 730
InterNIC (InterNetwork Information Center) 730
intrusion detection services 182, 797
intrusion detection services (IDS)
  IDS policy 802
IP (internet protocol), definition 4
IP address selection, source 244
IP addressing, virtual 361
IP security 883
IPCONFIG
  ARPTO 240
  DATAGRAMFWD 241, 418
  DYNAMICXCF 238, 241, 255, 385
  IGNOREREDIRECTS 284
  MULTIPATH 241, 288
  PATHMTUDISC 241, 284
  SOURCEVIPA 241, 256, 365
  SYSPLEXROUTING 241, 418
IPCONFIG6
  DYNAMICXCF 256
IPSec, security 154
IPv6
  autoconfiguration, stateless 258
  BPXPRMxx, sample definitions 39
IPv6 *(continued)*
  configuring static VIPAs 365
  defining TCP/IP as UNIX System Services PFS 39
  duplicate address detection 258
  dynamic routing 291
  InetD configuration file, setting up 1327
  IP security considerations 898
  OSPF security considerations 899
  router advertisements 357
  stack functions supported 11
  static routing 285
  static versus dynamic routing 281
iQDIO 6
ISMv2 540
IUCV/VMCF 120

J

JES 713
JESGETBYDSN 708
JESINTERFACELEVEL 708
JESLRECL 708
JESPUTGETTO 708
JESRECFM 708

K

Kerberos 662
Kerberos, security 161
key generation commands 1242
keyboard 1429

L

LAN channel station (LCS), general description 5
LCS (LAN channel station), general description 5
LDAP server
  Object classes 1381
  Schema definition 1388
LFS (logical file system), general description 7
license, patent, and copyright information 1433
link groups, SMC-R 512
links, SMC-D 513
links, SMC-R 510
Load Balancing Advisor (z/OS)
  overview 1147
load libraries, protecting with RACF 39
local host table 264
log files, offloading 228
logical file system (LFS), general description 7
LPD
  banner page 1271
  configuration 1269
  configuration data set 1271
  description 1269
  LPDDATA 1270
  LPDPRFX 1270
  PROFILE.TCPIP changes 1269
  tracing 1270
LRECL 677
LU assignments - objects, client identifiers, mapping statements 607

LU name groups, shared 586

M

main route table
  definition 279
mainframe
  education xliv
MAKESITE 265
management information base (MIB), general description 1237
MD5
  and OSPF 301
memory buffer 514
message catalogs, customizing 272
message data sets, customizing 276
messages data sets 276
messages, logging of 30
messages, TCP/IP
  rules for customizing 277
MGMTCLASS 677
MIB (management information base), general description 1237
MIBS.DATA 1257
middle-level qualifier (MLQ) 14
MIGRATEVOL 677
MISC server
  configuring 1324
  description 1323
  protocols supported 1323
  specifying server parameters 1325
  tracing 1325
MLQ (middle-level qualifier) 14
MODDVIPA utility 376
MODDVIPA, defining RACF profile for 377
MODIFY command
  Remote Execution server 1311
monitoring DNS name server responsiveness
  overview 761
monitoring network interfaces 451
MPC (multipath channel) 5
MPCPTP (multi-path channel point-to-point), general description 6
multi-path channel point-to-point, general description 6
multilevel secure environment
  overview 191
  required configuration 195
MULTIPATH (IPCONFIG MULTIPATH) 241, 288
multipath channel, general description 5
multiple application-instance scenario 372
multiple copies of TCP/IP 45
multiple stacks
  AUTOLOG 262
  BPXPRMxx 54
  CINET PFS 45
  configuring FTP with 675
  generic versus specific affinity 46
  OSA/SF considerations 1264
  OSPF and RIP considerations 294
  overview 45
  port management 46
  selecting a stack 51
  socket application programs 51
  TCPIP.DATA 51, 234
multiple stacks *(continued)*
  VIPA considerations 363, 370
MVS
  accounting 32
  automatic restart manager (ARM) 29
  component trace 114
  failure management 416
  general description 3
  logging system messages 30
  SERVAUTH 34
  system symbols 28, 236, 237

N

name resolution
  HOMETEST command to verify 270
  iterative resolution 731
  TESTSITE command to verify 270
  using HOSTS.LOCAL data set 264
  VIPA host 366
name servers
  authoritative 730
  caching-only, definition 732
  for VIPA host-name resolution 366
  forwarder, definition 732
  master, definition 732
  secondary, definition 732
  Stealth, definition 733
naming conventions, dynamically allocated data sets 14
NCS interface
  configuration 1280
  LLBD cataloged procedure 1280
  NRGLBD cataloged procedure 1280
  specifying statements in PROFILE.TCPIP 1281
NETSTAT 116
Netstat access control 145
NetView 1237, 1260
Network access control 140
network file system, see also NFS 1273
network interfaces monitoring 451
network management application 1238
network protocol layer, z/OS Communications Server TCP/IP 6
Network SLAPM2 subagent 843, 1241
NFS (network file system)
  PORTMAP address space 1273
nslookup command
  overview 735

O

offloading log files 228
OMPROUTE
  autolog considerations 298
  cataloged procedure 298
  configuring 296
  displaying information 326
  interaction with service policy 295
  interaction with VIPA 241, 295, 363
  multiple stack considerations 294
  overview 288, 294
  parameters 302
  ROUTESA_CONFIG 1246

OMPROUTE *(continued)*
run-time environment 293
sample configuration files 347
SNMP subagent 1261
starting 302
stopping 304
subagent 1241
supported protocols 289
verification of configuration and state 326
OMPROUTE_CTRACE_MEMBER 301
OMPROUTE_DEBUG_FILE 300
OMPROUTE_DEBUG_FILE_CONTROL 300
OMPROUTE_FILE 300
OMPROUTE_IPV6_DEBUG_FILE 300
OMVS RACF segment 36, 116
onslookup command
command line mode 736
interactive mode 736
overview 735
open shortest path first, see also OSPF 280
Open Systems Adapter (OSA)
with ARP offload 257
with SNMP 1261
OSA 549
OSA routing 61
OSA-Express feature
network traffic analyzer trace 104
VMAC routing 61
OSA-Express2 or later feature
synchronization of diagnostic data 105
OSNMPD, configuring 1245
OSNMPD.CONF, search order for 18
OSNMPD.DATA, search order for 19
OSPF (open shortest path first)
configuring authentication 301
configuring OSPF and RIP 305
definition 280
IPv6 280
overview 289
sample configuration files 347
security 162
otelnetd 659
outbound 45
outbound serialization 45

P

parameter, Subnet_mask 309
parameters, LPD server cataloged procedure
DIAG 1270
LPDDATA 1270
LPDPRFX 1270
TRACE 1270
TYPE 1270
VERSION 1270
parameters, Miscellaneous server
CHARGEN 1325
DEbug 1325
DISCARD 1325
ECHO 1325
TRACE 1325
Pascal sockets
general description 8
path length 42
PATHMTUDISC (IPCONFIG PATHMTUDISC) 241, 284
PCIe 504
PDSTYPE 677
performance considerations 41
PFID 552
PFS (physical file system) 7, 39, 45
physical file system (PFS) 45
physical file system, general description 7
policies
IDS 864
sysplex distributor 833
Policies
Attack 855
defining using LDAP 1389
Differentiated Services (DS) 831
DS 839
IDS Attack 869
IDS Scan 867
IDS TR 876
IDS TR TCP 861
IDS TR UDP 862
in Policy Agent configuration file 838
Integrated Services (RSVP) 833
RSVP 840
RSVP in LDAP 1393
Scan 851
sysplex distributor 840
sysplex distributor in LDAP 1395
Traffic Regulation (TR) 861
Policy Agent
and LDAP objects 1386
components 787
Configuration file 838
Configuring 808
overview 787
roles 787
sample files 800
sample LDAP objects, using 1388
Starting and stopping 823
types 787
policy-based route table
definition 279
policy-based routing
definition 280
port access control 137
port management
multiple stacks 46
port selection 247
port selection, ephemeral 246
PORTMAP
cataloged procedure 1274
configuring 1273
ETC.RPC 1274
required by NFS 1273
security server 1274
starting 1275
PORTMAP address space
configuring 1273, 1275
starting PORTMAP 1275, 1276
updating the PORTMAP cataloged procedure 1274, 1276
PortMapper, z/OS UNIX
configuring 1275
PORTRANGE

PORTRANGE *(continued)*
TCP/IP profile statements 264
POSIX standard
application behavior in z/OS Communications Server 7
using z/OS UNIX sockets API with 9
prerequisite information xliv
PRIMARY 677
printer support, 328x 594
printf function 276
problem detection and recovery, sysplex 452
procedures, TCP/IP
MISCSERV (MISCSERV) 1324
RXSERVE (RXPROC) 1309, 1313
SNMPD (SNMPDPRC) 1255
SNMPQE (SNMPPROC) 1258
PROFILE.TCPIP
ARPAGE 240
ARPTO 241
AUTOLOG 262
BEGINROUTES 257
BSDROUTINGPARMS 241
changes needed for FTP 664
CLAWUSEDOUBLENOP 241
CONNECTINITINTERVAL 243
CONNECTTIMEOUT 243
DATAGRAMFWD 241
DATASETPREFIX 14
DELAYACKS 243
DYNAMICXCF 241
ECSALIMIT 241
EXPLICITBINDPORTRANGE 391
FINWAIT2TIME 242
FRRTHRESHOLD 243
GLOBALCONFIG 240, 391
HOME 256
IGNOREREDIRECT 241
INTERFACE 256
INTERVAL 242
IPCONFIG 241
IPCONFIG6 241
IPSECURITY 241
KEEPALIVEPROBEINTERVAL 242
KEEPALIVEPROBES 242
LINK 254
MAXIMUMRETRANSMITTIME 243
MULTIPATH 241
MVS system symbols 28
NONAGLE 243
NOUDPCHKSUM 243
PATHMTUDISCOVERY 241
physical characteristics, setting up 250
PING 269
POOLLIMIT 241
PORT 50, 262, 664, 1263
PRIMARYINTERFACE 256
QDIOACCELERATOR 241
QUEUEDRTT 243
REASSEMBLYTIMEOUT 241
reserved port number definitions, setting up 259
RESTRICTLOWPORTS 243, 244
RETRANSMITATTEMPTS 243
SACONFIG 1261, 1263
sample 250
PROFILE.TCPIP *(continued)*
search order 23, 237
SENDGARBAGE 243
SOMAXCON 241
SOURCEVIPA 241, 256
SRCIP 241
STOPONCLAWERROR 241
SYSPLEXROUTING 241
TCP/IP operating characteristics, setting up 238
TCPCONFIG 242, 243, 664
TCPMAXRCVBUFRSIZE 243, 664
TCPMAXSENDBUFRSIZE 243
TCPRCVBUFRSIZE 243
TCPSENDBFRSIZE 243
TCPTIMESTAMP 243
TIMEWAITINTERVAL 242
TRACERTE 269
TRANSLATE 256
TTLS 243
UDPCONFIG 243
UDPQUEUELIMIT 244
UDPRCVBUFRSIZE 244
UDPSENDBFRSIZE 244
verifying 269
verifying your configuration 268
VIPADYNAMIC 256
PROFILE.TCPIP, specifying configuration statements
EZAFTSRV 665
PORTMAP 1273, 1275
TCPIP 237
PROFINE.TCPIP
DEVICE 254
program control 38
program directory 114
protocol suite 30
protocol suite, z/OS Communications Server TCP/IP 4
pwtokey 1251

Q

QoS, see Quality of service (QoS) 831
Quality of service (QoS)
and Policy Agent 833
Quality of Service (QoS)
QoS Policy 802

R

RACF (Resource Access Control Facility)
authorizing sources 36
considerations for FTP server 666
considerations for REXEC server 1310
port access control 137
resource protection 124
REXEC access to MVS 1310
stack access control 137
starting OMPROUTE 299
user access control 125
RACF profile, defining for MODDVIPA 377
RAW protocol, general description 7
RDMA network interface card 504
RDMA over Converged Ethernet 504
REASSEMBLYTIMEOUT 241

receive window, TCP 43
RECFM 677
Remote Direct Memory Access 503
remote hosts, accessing using Telnet 569
remote memory buffer 514
rendezvous processing 507
RESOLVE_VIA_LOOKUP 264
resolver
  address space
    defining 745
    managing 747
    overview 745
  API calls 737
  applying interim fix 748
  cache
    deleting entries 760
    displaying contents 760
    migrating to resolver caching 761
    reordering 754, 756
    sorting 756
  cache reordering
    disabling for applications 758
  caching
    configuring (optional) 757
    description of cached information 751
    eliminating for some users 758
    managing cache 759
    organization of cached information 753
    overview 750
  creating a resolver setup file 742
  customizing
    DEFAULTTCPIPDATA statement 741
    global TCPIP.DATA file 741
    GLOBALTCPIPDATA statement 741
    overview 738
    setup file 738
    setup file processing 740
  default settings 738
  DNS name server responsiveness
    diagnosis 764
    notifications for 763
    optimal UNRESPONSIVETHRESHOLD setting 768
  examples of DNS name server monitoring 766
  Extension Mechanisms for DNS standards (EDNS0) 770
  functions
    overview 750
  global TCPIP.DATA file 741
  GLOBALTCPIPDATA statement 741
  manually restarting 748
  modifying the UNRESPONSIVETHRESHOLD value 769
  monitoring DNS name server responsiveness 761
  overview 729
  setup file 738
  setup file processing 740
  starting 737
resolver configuration files
  for host names outside local area 264
  MVS versus z/OS UNIX resolver 773
  overview 771
  search order 771
  setting environment variables 775
  TCPIP.DATA 234
  use with OMPROUTE 298
RESOLVER_CONFIG
RESOLVER_CONFIG *(continued)*
  overview 775
  pointing to TCPIP.DATA 235
  use by OMPROUTE 300
  when running multiple TCP/IP stacks 53
RESOLVER_IPNODES 775
resolvers, configuring host
  nslookup considerations 736
resolving conflicts, VIPA 399
Resource Access Control Facility, see also RACF 36
Resource Access Control Facility, z/OS UNIX security and, see also RACF 36
RETPD 677
REXECD
  cataloged procedure 1311
  configuring PROFILE.TCPIP 1310
  security considerations 1310
  UNIX 1309
  user exits 1311
  userid.RHOSTS.DATA 1311
REXECD, z/OS UNIX
  configuring InetD 1327
  considerations in CINET environment 49
  files 1313
  installation 1313
REXX sockets 8
RFC (request for comments)
  accessing online xlvi
RIP (Routing Information Protocol)
  configuring 305
  definition 280, 290
  interaction with VIPA 432
  reserving RIP UDP port for OMPROUTE 297
  sample configuration files 347
RMB 514
RNIC 504
RoCE 504
RoCEv2 547, 549
router advertisements 357
router, definition 280
routing
  daemons 280, 288
  definition 280
  dynamic VIPAs 431
  IGNOREREDIRECTS 284
  IPv6 dynamic 291
  IPv6 static 285
  MULTIPATH 288
  network design considerations 324, 325
  PATHMTUDISC 284
  SOURCEVIPA 365
  static versus dynamic 281
  verification of 358
Routing Information Protocol, see also RIP 280
rpcbind address space
  configuring 1276
RPCINFO 1274
RSHD 37
RSHD, z/OS UNIX
  configuring InetD 1327
  considerations in CINET environment 49
  files 1314
  installation exit 1314
RSVP

RSVP *(continued)*
Configuring 842
Policies 833
Starting and stopping 842

S

search order
configuration files 13
DATASETPREFIX value 14
ETC.IPNODES 780, 784
ETC.PROTO 16, 781, 785
ETC.SERVICES 17, 781, 785
FTP.DATA 17, 676
high-level qualifier (HLQ) 14
HOSTS.ADDRINFO 780, 784
HOSTS.SITEINFO 779, 783
LPD configuration file 1270
MIBS.DATA 1257
middle-level qualifier (MLQ) 14
OSNMPD.CONF 18
OSNMPD.DATA 19
overview 13
PAGENT.CONF 19
PROFILE.TCP/IP 23
PROFILE.TCPIP 19
PW.SRC 20
resolver configuration files 771
RSVPD.CONF 20
SNMPD.BOOTS 21
SNMPD.CONF 21
SNMPTRAP.DEST 22
STANDARD.TCPXLBIN 778, 782
TCPIP.DATA 24
TCPXLBIN data set 682
TRAPFWD.CONF 23
with DD cards in TCP/IP startup procedure 24
without DD cards in TCP/IP startup procedure 24
SECONDARY 677
Secure Shell (SSH), security 162
Secure Socket Layer, see SSL 1329
security
application 123
event reporting 182
Express Logon Feature (ELF) 160
FTP server 666
IPSec 154
multilevel 191
overview 123
principals 151
protecting data in the network 151, 154
protocols 154
RACF 36
resource protection 124
SSL and TLS 157
z/OS UNIX considerations 36, 38
security considerations 1299
send window, TCP 44
SERVAUTH
MVS considerations 34
resource protection 124
restricting access to port numbers by applications 34
restricting access to TSO and UNIX shell Netstat command 35
SERVAUTH *(continued)*
setting up 264
SERVAUTH class profiles
EZB.BINDDVIPARANGE.sysname.tcpname 401
EZB.CIMPROV.sysname.tcpname 148
EZB.FRCAACCESS.sysname.tcpname 146
EZB.FTP.sysname.ftpdaemonname.ACCESS.HFS 666
EZB.INITSTACK.sysname.tcpname 146
EZB.IPSECCMD.sysname.tcpprocname.* 1362, 1363
EZB.MODDVIPA.sysname.tcpname 377
EZB.NETACCESS.sysname.tcpname.zonename 140
EZB.NETMGMT.sysname.tcpname.IPSEC.DISPLAY 148
EZB.NETMGMT.sysname.tcpname.SYSTCPCN 146
EZB.NETMGMT.sysname.tcpname.SYSTCPDA 146
EZB.NETMGMT.sysname.tcpname.SYSTCPER 147
EZB.NETMGMT.sysname.tcpname.SYSTCPSM 147
EZB.NETSTAT.sysname.tcpname.netstat_option 145
EZB.OSM.sysname.tcpname 142
EZB.PAGENT.sysname.image.ptype 810
EZB.PORTACCESS.sysname.tcpname.port_safname 137
EZB.SNMPAGENT.sysname.tcpprocname 1252
EZB.SOCKOPT.sysname.tcpname.socketoption 143, 144
EZB.STACKACCESS.sysname.tcpname 137
SESSLIM parameter, VTAMLST 120
shared LU name groups 586
Shared Memory Communications - Direct Memory Access, *See* SMC-D
Shared Memory Communications multiple IP subnet support
SMCv2 534
Shared Memory Communications over Remote Direct Memory Access, *See* SMC-R
shortcut keys 1429
Simple Network Time Protocol (SNTP) 1305
SIOCSVIPA ioctl 375
SIOCSVIPA6 ioctl 375
SMC
concepts 507
displaying information 560, 563
interaction with
IDS 556
security functions 556
sysplex distributor 555
TCP application data transfer options 558
TCP keepalive 557
link groups 510
links 510
memory buffer 514
monitoring
SNMP 563
TCP/IP callable NMI 562
overview 503
physical network considerations 520
terms 515
type 540
version 540
VLANID 518
SMC filters 515, 539
SMC-D
configuring, steps for 532
direct memory buffer 515
DMB 515
ISM interfaces, managing 560
links 513

SMC-D *(continued)*
overview 503, 507
physical network considerations 520
preparing to use 532
rendezvous processing 507
stopping 565
system requirements 529
VLANID 518
SMC-Dv2
EID Format 537
enabling SMC-Dv2 540
ISM VCHID Association 541
Multiple EID 537
PNetID 541
SEID 538
SMC Enterprise ID 534
System EID 538
UEID 537
User Defined EID 537
using SEID 538
using UEID 538
SMC-R
10 GbE RoCE Express interfaces, managing 559
configuring, steps for 530
high availability 522
IBM 10 GbE RoCE Express feature 504
interaction with
MTU 558
packet trace 558
link groups 512
links 510
monitoring
SMF 563
network requirements 528
overview 503
PCIe 504
physical network considerations 520
preparing to use 529
remote memory buffer 514
rendezvous processing 507
RMB 514
RoCE 504
staging buffers 515
stopping 564
system requirements 527
VLANID 518
SMC-Rv1
migration 552
SMC-Rv2
link group 552
SMC-Rv2 resilience
high availability with RoCEv2 and IP routing 552
SMCRIPADDR 552
SMF (System Management Facility)
records for FTP 32, 685
records for Telnet 32
see also, accounting 32
user exit for FTP server 708
SMS (Storage Management System) 679
SNMP (Simple Network Management Protocol)
agent 1239
agents and subagents 1245, 1261
CINET 1254
community names 1247
SNMP (Simple Network Management Protocol) *(continued)*
community-based security 1246, 1258
configuring 1237
configuring for z/OS UNIX 1237
creating user keys 1251
DD statement for external message data set 276
Enterprise-Specific variables 1259
MIBDESC.DATA 1258
multiple SNMPv3 agents in same MVS image 21, 1250
NetView 1258
OSA 1261
OSNMPD, starting 1255
OSNMPD.DATA 19
overview 1237
port specification 1246
protocols 1238
PW.SRC 20
pwtokey 1251
security 1243
security models, overview 1239
SLAPM2 subagent 798
snmp 1256
SNMPD.BOOTS 1250
SNMPD.CONF 1250
SNMPTRAP.DEST 22, 1248
SNMPv1, SNMPv2C, SNMPv3 1243
subagents 1240
TCP/IP profile statements 264, 1245, 1263
textual names 1257
trap forwarding 1265
TRAPFWD.CONF 1265
updating the SNMPD cataloged procedure 1258
user-based security 1249
snmp command, configuring 1256
SNMP Network SLAPM2 subagent 798
snmp, general description 1238
SNMPIUCV module 1260
SNMPv3, security 162
SNTPD daemon 1305
socket APIs 8
socket applications (z/OS UNIX), support for fast path 42
socket applications use of z/OS UNIX 36
Sockets Extended
definition of call instruction API 8
definition of macro API 8
softcopy information xliv
source IP address selection 244
SOURCEVIPA (IPCONFIG SOURCEVIPA) 241, 256, 365
SPACETYPE 677
specific stack affinity 46
SPREAD 708
SQL 720
SQL usage
in FTP server 720
SQLCOL 708
SSL
for DCAS 1329
for FTP server 1329
for Telnet server 1329
overview 1329
security 157
stack access control 137
stack communications, z/OS UNIX to TCP/IP 42
stack functions supported, IPv6 11

stack, TCP/IP 3
STANDARD.TCPXLBIN 778, 782
START command 121
started task 36
static routes
  definition 280
static routing
  configuration examples 286
  IPv4 283
  IPv6 285
  versus dynamic routing 281
Storage Management System (SMS) 679
STORCLASS 677
subagent, Network SLAPM2 843
subagent, Network SLAPM2 subagent (nslapm2) 1241
subagent, OMPROUTE 1241
subagent, TCP/IP 1240
subagent, TN3270E Telnet 1241
Subnet_mask parameter 309
subnets, dynamic VIPAs within 415
subplexing, sysplex 436
summary of changes xlix, lii, liv
superuser authorization 36
SURROGATE, in anonymous logins 715
symbols, MVS system 28
syntax diagram, how to read xxxix, xlii
SYSFTPD 676
syslogd
  command format 221
  configuring 213
  diagnosing configuration problems 233
  ECSA storage mapping 234
  exit values 224
  for z/OS UNIX applications 231
  modes of operation 220
  overview 30
  remote messages, receipt of 225
  starting 219
  stopping 224
  token name 234
  usage notes 232
sysplex
  failure management 416
  network interfaces monitoring 451
  problem detection and recovery 452
  subplexing 436
  sysplex-wide dynamic source VIPAs for TCP connections 388
  Sysplex-Wide Security Associations (SWSA) 396
  SYSPLEXPORTS 389
  TCP/IP in a 435
  workload balancing 468
sysplex distributor
  configuring distributed DVIPAs 382
  DATAGRAMFWD 418
  dynamic port assignment 388
  policies 833
  policy interactions 475
  SYSPLEXROUTING 418
  timed affinities 393
  using IPSec with 398
  with Cisco routers 475
  with SWSA 398
  workload verification 427
Sysplex Distributor
  DYNAMICXCF 238
sysplex-wide dynamic source VIPAs for TCP connections 388
Sysplex-Wide Security Associations (SWSA)
  DVIPA takeover 396
  EZBDVIPAvvtt 396, 399
  IPSec with DVIPAs and sysplex distributor 398
  overview 396
  sysplex distributor 398
SYSPLEXPORTS 389
SYSPLEXROUTING (IPCONFIG SYSPLEXROUTING) 241, 418
SYSTCPD DD
  fork() considerations 234, 778
  search order for TCPIP.DATA 235
system and network requirements
  SMC 527
system symbols, MVS 28, 236

T

takeover, DVIPA 396
tasks
  (VTAM configuration data set, customizing)
    steps 573
  ADNR in a sysplex subplexing environment, using
    steps 1215
  ADNR, configuring
    steps 1197
  ADNR, granting authority to start
    steps 1201
  Advisor, granting authority to start
    steps 1150
  Agents, granting authority to start
    steps 1151
  applying an interim fix
    steps 748
  AT-TLS, starting and verifying its operation
    steps 1130
  authorizing resources for NSS
    steps 1086
  automated domain name registration, configuring
    steps 1197
  avoiding adjacency failures
    steps for 204
  branch office model: part 1 (host-to-gateway with IPSec), configuring
    steps 1017
  branch office model: part 2 (gateway-to-gateway with IPSec), configuring
    steps 1032
  branch office with NAT model (host-to-gateway with IPSec), configuring
    steps 1026
  bridge, creating and customizing
    sendmail 1298
    steps 1298
  Configuring OMPROUTE
    steps for 296
  configuring OSPF and RIP (IPv4 and IPv6)
    steps for 305
  configuring static VIPAs for a z/OS TCP/IP stack
    steps for 365

tasks *(continued)*
creating a resolver setup file (optional)
steps 742
creating a separate home directory for each security label
steps for 200
creating a separate resolver configuration file for each security label
steps for 201
CSFSERV resource class, setting up profiles in
steps 1365
CSSMTP, configuring and starting
steps 1286
CSSMTP, configuring SMF records for
steps 1295
CSSMTP, creating mail on the JES spool data set for
steps 1287
CSSMTP, granting authority to start
steps 1291
CSSMTP, using Transport Layer Security for
steps 1293
customizing the FTP client for Kerberos
steps for 700
customizing the FTP client for TLS
steps for 695
customizing the FTP server for Kerberos
steps for 691
customizing the FTP server for TLS
steps for 686
defining the resolver address space
steps 745
DMD, authorizing resources for
steps 1118
DMD, configuring
steps 1116
DVIPA within a specific VIPARANGE subnet, controlling creation by SIOCSVIPA ioctl or MODDVIPA utility
steps 379
DVIPA within a specific VIPARANGE subnet, controlling whether an application can bind to create
steps 403
DVIPA, controlling creation by SIOCSVIPA ioctl or MODDVIPA utility
steps 378
DVIPA, controlling which applications can bind to create
steps 402
enabling Policy Agent load distribution functions
steps for 486
existing key database, migrating to a RACF key ring
steps 1376
FTP server access, controlling
steps 668
FTP server, assigning password phrases to user IDs
steps 672
FTP server, setting up port of entry for users
steps 669
FTP server, setting up security
steps 667
FTPD cataloged procedure, configuring
steps 665
global definitions for all stacks, configuring
steps 197
tasks *(continued)*
hot standby distribution, configuring
steps 482
IKE daemon, authorizing to RACF
steps 1362
IKE daemon, configuring
steps 1045
IKE daemon, preparing to run
steps 1361
IKE daemon, setting up for RSA signature mode authentication
steps 1370, 1372
interface, configuring for the intraensemble data network (CHPID type OSX)
steps 494
intraensemble data network, enabling HiperSockets access to
steps 498
intranode management network (CHPID type OSM), enabling IPv6 on a stack for access to
steps 499
intranode management network (CHPID type OSM), using
steps 499
IP security policy, configuring
overview 939
steps 972
IP security policy, local, configuring using both a stack-specific file and a common file
steps 945
IP security policy, local, configuring using only a common IP security configuration file
steps 942
IP security policy, local, configuring using only a stack-specific IP security configuration file
steps 943
IP security policy, remote, configuring using both a stack-specific file and a common file
steps 945
IP security policy, remote, configuring using only a common IP security configuration file
steps 942
IP security policy, remote, configuring using only a stack-specific IP security configuration file
steps 944
IP security, using
overview 888
ipsec command, authorizing resources for
steps 1118
ipsec command, authorizing to RACF
steps 1363
JES, setup
steps 1288
Load Balancing Advisor (z/OS), configuring
steps 1152
Load Balancing Advisor (z/OS), preparing to use
steps 1148
Load Balancing Advisor (z/OS), starting
steps 1171
LUNR, defining
steps 588

tasks *(continued)*
LUNS, defining
steps 588
manually restarting the resolver 748
message catalog, creating a modified
steps 275
message catalog, creating from the shipped catalog and preserving its timestamp
steps 683
migrating the FTP server and client to use AT-TLS
steps for 703
migrating to resolver caching
steps 761
modifying the UNRESPONSIVETHRESHOLD value
steps 769
modifying UNRESPONSIVETHRESHOLD value
optimal setting 768
NCS interface, configuring
steps 1280
NSS server, configuring
steps 1095
partner company model (host-to-host with IPSec), configuring
steps 985
partner company with NAPT model (host-to-host with IPSec), configuring
steps 1012
partner company with NAT model (host-to-host with IPSec), configuring
steps 998
Policy Agent, configuring
steps 808
policy changes, managing
steps 791
PORTMAP address space, configuring
steps 1273
profiles to control access to the RACDCERT command, defining
steps 1371
QDIO inbound workload queueing, enabling
steps 73
RACF facilities and access controls, defining
steps 1370
resolver address space
overview 745
resolver cache, displaying contents
step 760
resolver cache, managing
steps 759
resolver, configuring caching (optional)
steps 757
resolver, deleting cache entries
steps 760
resolver, disabling cache reordering for applications
steps 758
resolver, eliminating caching for selected users
steps 758
resolver, managing the cache
steps 759
resource profile, defining with RACF
steps 1152
rpcbind, configuring
steps 1276
tasks *(continued)*
security for the procedure name and the associated user ID, defining
steps 571
self-signed X509 digital certificate for the IKE daemon, generating
steps 1375
SERVAUTH class, activating and defining
steps 667
setting stack affinity by security label
steps for 201
SMTPNOTE CLIST, customizing
steps 1289
starting SNTPD as a procedure
steps for 1306
starting SNTPD from the z/OS shell
steps for 1305
subplex, partitioning a set of TCP/IP stacks in a sysplex into a
steps 439
subplexing, preparing your sysplex for
steps 438
syslogd, configuring archive details
steps 231
syslogd, configuring archive triggers
steps 230
TCP/IP profile, converting from IPv4 IPAQENET DEVICE, LINK, and HOME definitions to the INTERFACE statement
steps 57
TCP/IP profile, converting from IPv4 IPAQIDIO DEVICE, LINK, and HOME definitions to the INTERFACE statement
steps 75
TCP/IP profile, converting from IPv4 VIRTUAL DEVICE, LINK, and HOME definitions to the INTERFACE statement
steps 367
TN3270E Telnet server configuration data set, customizing
steps 575
trusted internal network model (simple IP filtering), configuring
steps 975
VLANs for Hipersockets CHPID, configuring
steps 80
X509 digital certificate, generating and having it signed by a certificate authority
steps 1373
z/OS Load Balancing Advisor, configuring
steps 1152
z/OS Load Balancing Advisor, preparing to use
steps 1148
z/OS Load Balancing Advisor, starting
steps 1171
z/OS system, preparing for IP security
steps 935
z/OS UNIX file system, controlling access to

tasks *(continued)*
z/OS UNIX file system, controlling access to *(continued)*
steps 670
z/OS UNIX PORTMAP address space, configuring
steps 1275
TCP (transmission control protocol), definition 4
TCP receive window 43
TCP send window 44
TCP/IP
application configuration files 25
configuration data sets 15
configuration files for the stack, search order 23
customizing messages 271
dynamic configuration changes 237
initialization failure 115
installation, planning for 113
multiple instances 45
online information xlvi
protocol specifications 1409
protocol suite 3
resolver configuration files 771
search order and configuration files for the stack 13
socket APIs 8
stack configuration files, search order 23
starting the address space 121
startup DD cards 24
sysplex considerations 435
TCP/IP configuration data sets 15
TCP/IP subagent 1240
TCPIP.DATA
/etc/resolve.conf 234
characteristics 235
configuring for FTP 675
creating 235
DATASETPREFIX 14, 53
finding with SYS1.TCPPARMS 15
multiple stacks 51
MVS system symbols 28, 236
overview 234
search order 24
syntax 235
TCPIPJOBNAME 53
verifying 269
TCPSTACKSOURCEVIPA 241, 364, 365, 388
Technotes xliv
Telnet configuration data set
customizing 575
updating with VARY TCPIP,tnproc,OBEYFILE command 579
TEMPIP 107
TESTSITE 270
TIMED daemon 1303
TLS
for DCAS 1329
for FTP server 1329
for Telnet server 1329
security 157
TN3270E Telnet server
accounting 32
associated printer function 627
configuration data set 574
connection and session takeover 637
connection security 597
TN3270E Telnet server *(continued)*
connection types 592
device types 643
diagnostics 580
disconnect on error 642
Express Logon Feature (ELF) 643
Generic connection requests 625
getting started 569
keep LU for client identifier 629
keeping the ACB open 642
logmode considerations 643
LU assignments 607
LU group capacity warning 629
LU mapping by application name 630
LU mapping selection rules 631
LU name assignment user exit 626
LUMAP statements, multiple 628
managing 577
map default application and ParmsGroup by LU group 628
mapping groups to client identifiers 625
mapping objects to client identifiers 607
multilevel security 634
Network Access Control 604
overview 569
queueing sessions 641
session initiation management 635
SMF records 648
solicitor 644
Specific connection requests 625
storage considerations 597
timers 604
Unformatted System Services (USS) 644
using wildcards to configure 610
TN3270E Telnet subagent 1241
TNF 117
trademark information 1436
translation of data, FTP 680
transmission control protocol (TCP), definition 4
transport layer, z/OS Communications Server TCP/IP 7
Trap Forwarder Daemon 1242
traps and SNMP, definition 1238
TRMD
running as a started task 880
running from the z/OS UNIX shell 881
stopping and starting 879
TRMDSTAT 881

U

UCOUNT 677
UDP (user datagram protocol), general description 7
UMASK 677
unique application-instance scenario 372, 373
UNITNAME 677
UNIX, superuser authorization 36
user datagram protocol (UDP), general description 7

V

variables, setting environment for resolver configuration files 775

VARY 1299
VCOUNT 677
verification
  system configuration 121
  X Window System 270
VIPA (virtual IP address)
  backing up TCP/IP stack 368
  configuring static 365
  distributed DVIPA 361
  dynamic (DVIPA) 369
  dynamic routing 361
  dynamic VIPA 361
  interfaces 312
  manual movement 363
  overview 55, 361
  static 364
  takeover planning 362, 370
VIPA interfaces 312
virtual IP address, see also VIPA 361
virtual MAC routing 61
virtual machine communication facility, see also VMCF 117
VLANID, and SMC 518
VLANID, and SMC-D 518
VLANID, and SMC-R 518
VMAC routing 61
VMCF (virtual machine communication facility)
  commands 118
  configuring as non-restartable system 118
  configuring as restartable system 117
VOLUME 677
VPN, security 154
VTAM parameters, general update 120
VTAM, online information xlvi
VTAMLST member 120

W

well-known procedure names, defining 1274

X

X Window System verification 270
X Window, verifying installation 270
XPG4 standard
  using z/OS UNIX sockets API with 9

Z

z/OS Basic Skills Information Center xliv
z/OS Communications Server environment, overview 13
z/OS Communications Server overview 3
z/OS Container Extensions
  configuring application-instance DVIPAs 380
z/OS Load Balancing Advisor
  overview 1147
z/OS Management Facility, IBM
  AT-TLS 1122
  DMD 1116
  IDS 863
  overview 789
z/OS Management Facility, IBM *(continued)*
  policy-based routing 350
  QoS 837
z/OS UNIX
  security considerations 38
z/OS UNIX initialization failure 116
z/OS UNIX sockets 9
z/OS UNIX System Services (z/OS UNIX)
  applications and syslogd 231
  concepts 11
  hierarchical file system concepts 12
  overview 11
z/OS UNIX Telnet server, configuring 655
z/OS UNIX, superuser authorization 36
z/OS, documentation library listing 1437
ZERT 805
zERT enforcement policy 176–179, 181, 182
zERT policy-based enforcement (ZERT) 805
zone transfers 732

Communicating your comments to IBM

Important: If your comment regards a technical question or problem, see instead “If you have a technical problem” on page 1457.

Submit your feedback by using the appropriate method for your type of comment or question:

Feedback on z/OS function
If your comment or question is about z/OS itself, submit a request through the IBM RFE Community (www.ibm.com/developerworks/rfe/).

Feedback on IBM Documentation function
If your comment or question is about the IBMDocumentation functionality, for example search capabilities or how to arrange the browser view, send a detailed email to IBM Documentation Support at ibmdocs@us.ibm.com.

Feedback on the z/OS product documentation and content
If your comment is about the information that is provided in the z/OS product documentation library, send a detailed email to mhvrcfs@us.ibm.com. We welcome any feedback that you have, including comments on the clarity, accuracy, or completeness of the information.

To help us better process your submission, include the following information:

- Your name, company/university/institution name, and email address
- The title and order name of the document, and the version of z/OS Communications Server
- The section title of the specific information to which your comment relates
- The text of your comment.

When you send comments to IBM, you grant IBM a nonexclusive authority to use or distribute the comments in any way appropriate without incurring any obligation to you.

IBM or any other organizations use the personal information that you supply to contact you only about the issues that you submit.

If you have a technical problem

If you have a technical problem or question, do not use the feedback methods that are provided for sending documentation comments. Instead, take one or more of the following actions:

- Go to the IBM Support Portal (support.ibm.com).
- Contact your IBM service representative.
- Call IBM technical support.

© Copyright IBM Corp. 2000, 2022

Product Number: 5650-ZOS

SC27-3650-50