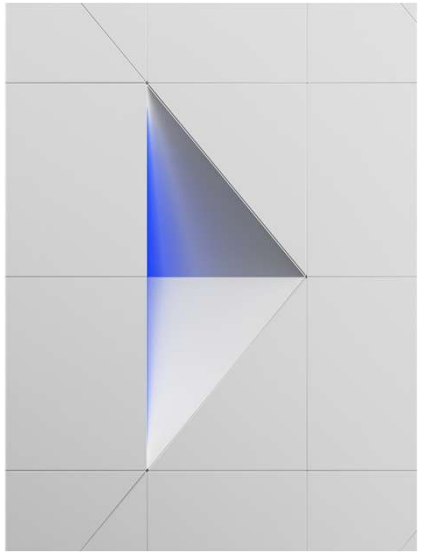
IBM Z Tech Bytes – May 2025

z/OS Communications Server use of Network Express Feature

Linda Harrison
lharriso@us.ibm.com

Contents



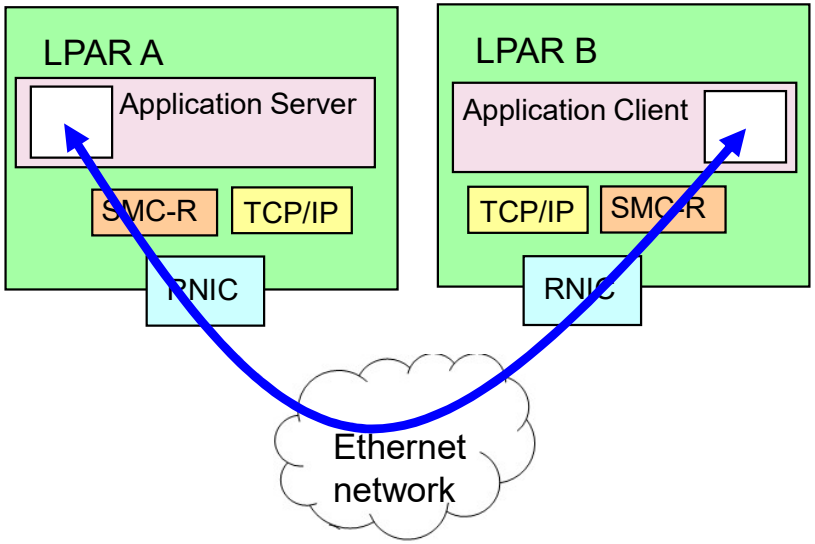
- Network Express Physical Adapter
- Protocols
- Interface Parameters
- Migration Option
- Display Output
- Limitations/Guidelines
- APARs
- Appendix – Backup Charts

Network Express Physical Adapter

OSA CHPID Type

- OSA > OSA-2 > OSA-Express > OSA-Express2 > … > OSA-Express7
- OSA CHPID type OSE - **obsolete**
  - Non-QDIO mode connection to external Ethernet LAN supporting both TCP/IP and SNA traffic concurrently
- OSA CHPID type OSX - **obsolete**
  - QDIO mode connection to Intraensemble Data Network (IEDN)
- OSA CHPID type OSM - **obsolete**
  - QDIO mode connection to Intranode Management Network (INMN)
- OSA CHPID type OSN - **obsolete**
  - OSA for NCP mode connects NCP running in LPARs with Communication Controller for Linux on System z (CCL) to LPARs running TPF and VTAM in z/OS, z/VM, and z/VSE.
- **OSA CHPID type OSD**
  - QDIO mode connection to external Ethernet LAN
- **OSA CHPID type OSC**
  - Console mode connection to external Ethernet LAN for console, TN3270, and telnet access to LPARs.

SMC-R and RoCE



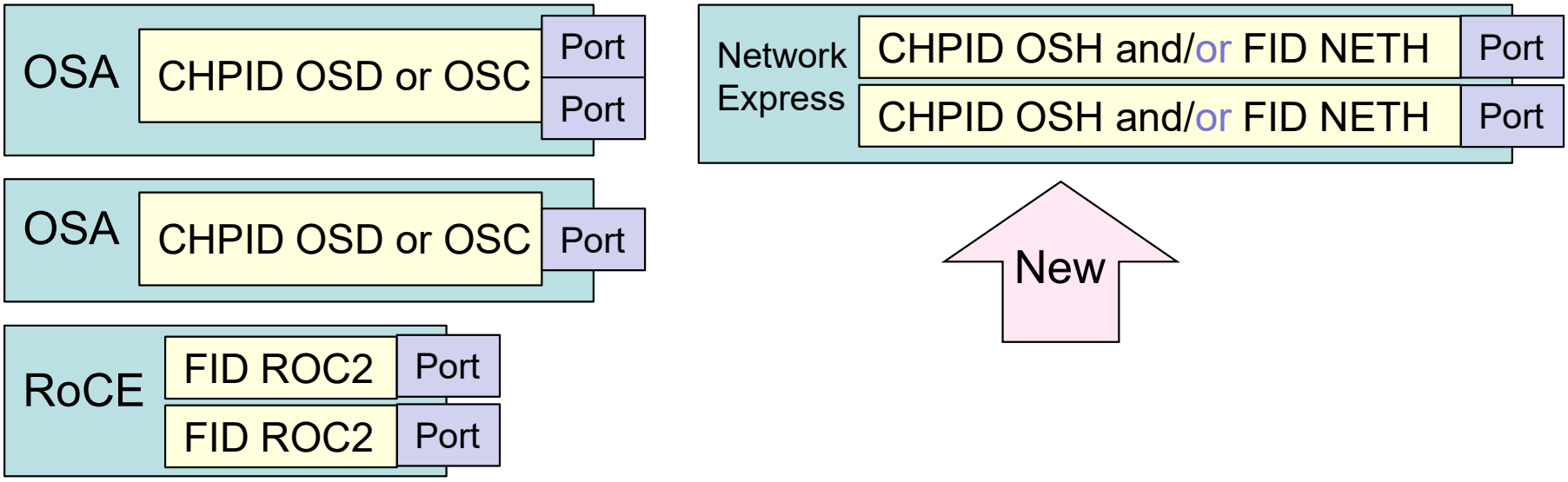
- Shared Memory Communication over RDMA (SMC-R)
  - Is a sockets over RDMA communication protocol that allows existing TCP applications to transparently benefit from RoCE.
  - Requires no application change.
  - Provides host-to-host direct memory access without the traditional TCP/IP processing overhead.
  - Allows customers to benefit from InfiniBand technology by leveraging their existing 10GbE Ethernet infrastructure.
  - TCP protocol only! No UDP (ie. EE), SNA, etc.
    - All TCP traffic except IPsec

- z/OS includes SMC-R support.
- SMC-R is only used over the RoCE Express feature to a partner SMC-R and RoCE.
  - While other platform (non-Z) RoCE RNICs might exist in your network, the Z RoCE Express feature is only able to communicate with them if they use SMC-R as well.

RoCE Feature

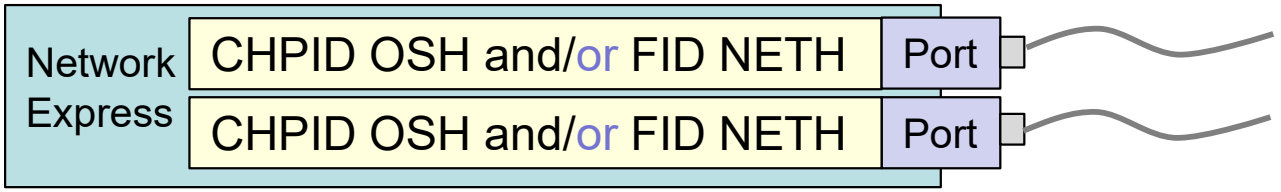
- RoCE Express
- RoCE Express2
- RoCE Express3
- Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE)
  - Linux supports the RoCE feature sending data using TCP/IP and Shared Memory Communications over RDMA (SMC-R).
  - z/OS Communications Server only supports RoCE feature sending data using SMC-R.
    - QDIO (OSD) OSA is required for sending TCP/IP data.
    - SMC-R and RoCE are required for sending data using RDMA.
    - In either case, the application is still written to communicate using TCP/IP.
  - RoCEv2 support allows multiple subnets.

New Network Express Feature



- The Network Express feature combines the functionality of both the OSA-Express in OSD (QDIO) mode and the RoCEv2 Express feature.
- Console support still requires the OSA-Express in OSC mode.

Protocols and Ports



- Network Express feature supports new protocols (H for Hybrid)
  - OSH protocol - Enhanced QDIO (EQDIO) — OSA QDIO like
  - NETH protocol - RoCE, RDMA, SMC-R, TCP/IP (Linux) — RoCE like
- Network Express physical feature
  - 2 CHPIDs per card - 2 Ports per card
    - 1 Port per CHPID
  - 10GbE or 25GbE
  - Both PCHIDs must be the same type
    - Can’t mix ports with different speed or optics
  - “OSA” name will still be used as well in manuals
- Network Express port support
  - CHPID OSH and FID NETH – one or both supported on the same port
  - LPAR to LPAR traffic supported (as OSA always has)

Protocols

EQDIO Protocol

- For each Network Express feature port
  - Define CHPID type OSH for EQDIO - OSA QDIO like support
  - Define FID type NETH for SMC-R - RoCE like support
- EQDIO Protocol
  - CHPID type OSH
  - A single Device Number
  - Control Queues replace OSD Read/Write devices
  - VTAM TRLE is dynamically created
  - New Interface EQENET
  - Layer 2
  - Default MTU is 9000 (Jumbo Frames)
- Linux Network Express does not use EQDIO since Linux RoCE support already includes SMC-R and **TCP/IP**.

NETH Protocol

	OSH Only	NETH Only	OSH and NETH
z/OS	YES	NO	YES
Linux	NO	YES	NO

- NETH Protocol
  - PFID type NETH
  - PF defined in OSA firmware
  - No more Management of RoCE PFIDs in Resource Groups
  - No more port parameter
    - Port number is NOT configured anywhere (HCD or the OS)
  - No z/OS Communications Server SMC-R configuration changes
- z/OS Communications Server Network Express Support
  - CHPID OSH, Interface EQENET **required**
  - PFID NETH **optional** – not supported without OSH defined
    - Interface EQENET SMCRIPADDR parameter required
    - NETH (Interface EZARIUTxyyyy) dynamically created/started when Interface EQENET is started
      - x is always 1 for port and yyyy is the PFID number

Interface Parameters

EQENET and EQENET6 Parameters

- Interface inf_name DEFINE **EQENET DEVNUM xxxx VMAC IPADDR ipaddr/mask**…
  - **DEVNUM** replaces PORTNAME – **required**
  - **IPADDR** – **required (subnet mask required for IPv4)**
  - **VMAC** – **will be auto-generated if missing**
  - Other parameters that remain valid
    - MTU
    - VLANID
    - SECCLASS
    - MONOSYSPLEX
    - ISOLATE
    - SOURCEVIPAINTERFACE
    - TEMPIP
    - SMCR
    - SMCD
  - Parameters not supported/needed
    - CHPIDTYPE
    - PORTNAME
    - PRIROUTER/SECROUTER
    - INBPERF (eliminated, internal capability is DYNAMIC with IWQ)
    - READSTORAGE (eliminated, dynamically managed)
    - DYNVLANREG
    - OLM – IPAQENET only
    - ADDADDR – IPAQENET6 only
    - DELADDR – IPAQENET6 only
    - ADDTEMPPREFIX – IPAQENET6 only
    - DELTEMPPREFIX – IPAQENET6 only

DEVNUM Parameter

- Device number (DEVNUM) allows a specific OSH CHPID to be identified where any available (arbitrary) device (under the specific OSH CHPID) can be selected and used.
  - Configure the **first device number** in the defined range, such as 2F00, for all INTERFACE statements for a given CHPID, and the first available device will be used.
- The configured device and the actual device selected for the INTERFACE are not always the same (OSH NETSTAT displays both the configured and actual device)
- If multiple INTERFACEs are configured for the same OSH CHPID on a TCP/IP stack, then each INTERFACE must conform to the following rules:
  - The configured DEVNUM value must be the same
  - A unique VLAN ID for each INTERFACE is required
  - VMAC with Route ALL setting is required
  - Maximum number of interfaces supported per IP protocol is 32

Migration Option

Migration

- As a **TEMPORARY** migration option from IPAQENET to EQENET, add the following to IPAQENET:
  - DEVNUM parameter
  - IPADDR with subnet mask (IPv4) if not already defined
  - VMAC if not already defined (if missing it will by dynamically created)
- If IPAQENET is started on a pre-z17 the Interface will be brought up as an IPAQENET, ignoring DEVNUM.
- If IPAQENET is started on a z17 the Interface will be brought up as an EQENET, ignoring PORTNAME and other IPAQENET only parameters.

Migration

- Temporary IPAQENET example:

```
INTERFACE O4ETHA2 DEFINE IPAQENET
DEVNUM 2B70
PORTNAME PORTQDIO
IPADDR 16.12.37.160/20
VLANID 702
MTU 1500
READSTORAGE GLOBAL
INBPERF DYNAMIC WORKLOADQ
SMCR PFID 3140 SMCRIPADDR 16.12.37.165
```

- Final EQENET example:

```
INTERFACE O4ETHA2 DEFINE EQENET
DEVNUM 2B70
IPADDR 16.12.37.160/20
VLANID 702
MTU 1500
SMCR PFID 3140 SMCRIPADDR 16.12.37.165
```

- Temporary IPAQENET6 example:

```
INTERFACE V6O4ETHA2 DEFINE IPAQENET6
DEVNUM 2B70
PORTNAME PRT6QDIO
INTFID 0:16:207:3
IPADDR 2001:0DB8:172::16:207:13
VLANID 601
MTU 9000
READSTORAGE GLOBAL INBPERF DYNAMIC WORKLOADQ
```

- Final EQENET6 example:

```
INTERFACE V6O4ETHA2 DEFINE EQENET6
DEVNUM 2B70
INTFID 0:16:207:3
IPADDR 2001:0DB8:172::16:207:13
VLANID 601
MTU 9000
```

Change Interface type to EQENET/6 after move to z17.

Display Output

Display CHPID

```
08.57.47 D M=CHP(09)
08.57.47 IEE174I 08.57.47 DISPLAY M 867
CHPID 09: TYPE=35  DESC=OSA HYBRID,  ONLINE
DEVICE STATUS FOR CHANNEL PATH 09
    0 1 2 3 4 5 6 7 8 9 A B C D E F
2B7 + + + + . . . . . . . . . . . .
SWITCH DEVICE NUMBER = NONE
******************** SYMBOL EXPLANATIONS ********************
+ ONLINE @ PATH NOT VALIDATED - OFFLINE . DOES NOT EXIST
* PHYSICALLY ONLINE $ PATH NOT OPERATIONAL
```

CHPID Type 35 for OSH

The number of devices defined in VTAM is not required to match the number in HCD.

The VTAM definition can’t exceed the number defined in HCD.

Display PCIE

```
09.38.57 D PCIE
09.38.57 IQP022I 09.38.57 DISPLAY PCIE 767
PCIE 0010 ACTIVE
PFID     DEVICE TYPE NAME STATUS ASID JOBNAME CHID VFN PN
00000501 Network Express  CNFG                 02FC 0001 1
00000502 Network Express  CNFG                 02FC 0002 1
```

Port Number is always 1.

```
09.58.32 D PCIE,PFID=501
09.58.32 IQP024I 09.58.32 DISPLAY PCIE 776
PCIE 0010 ACTIVE
PFID     DEVICE TYPE NAME STATUS ASID JOBNAME CHID VFN PN
00000501 Network Express  CNFG                 02FC 0001 1
CLIENT ASIDS: NONE
PNetID 1: ZOSNET
```

Migration Netstat DevLinks

```
INTFNAME: O4ETHA2 INTFTYPE: EQENET INTFSTATUS: READY
**AUTOMIGRATED**
DEVNUM: 8102 ACTDEVNUM: 8104 DEVSTATUS: READY
CHPIDTYPE: OSH CHPID: 71 PCHID: 0139
SMCD: YES SMCR: YES PNETID: NETWORK7CD
ADAPTER GEN: NETWORK EXPRESS V1.0
TRLE: IUTE8104 CODE LEVEL: 303100000000001440
PORTNAME: EZAP8104
SPEED: 0000010000 (10G)
IPBROADCASTCAPABILITY: NO
VMACADDR: 42000C69C925 VMACORIGIN: OSA VMACROUTER: ALL
CFGMTU: 9000 ACTMTU: 9000
IPADDR: 16.12.37.160/20
VLANID: 702 VLANPRIORITY: DISABLED
TOTAL READ STORAGE: 19.0M
CHECKSUMOFFLOAD: YES SEGMENTATIONOFFLOAD: YES
SECCLASS: 255 MONSYSPLEX: YES
ISOLATE: NO
```

AUTOMIGRATED indicates that IPAQENET was automatically migrated to EQENET.
EQENET missing VMAC parameter is auto-generated.
IP Address and Subnet Mask is required.

Migration Netstat DevLinks (cont.)

```
DISPLAY OSAINFO RESULTS FOR INTFNAME: O7ETHB0
DATAPATH: 2FA0 REALADDR: 0045
PCHID: 0114 CHPID: D2 CHPID TYPE: OS?
OSA CODE LEVEL: 3031240022221484
ACTIVE SPEED: 10 GB/SEC GEN: NETWORK EXPRESS V1.0
MEDIA: MULTIMODE FIBER JUMBO FRAMES: YES ISOLATE: NO
PHYSICALMACADDR: 9C63C0530F12 LOCALLYCFGMACADDR: 9C63C0530F12
QUEUES DEFINED OUT: 5 IN: 8 ANCILLARY QUEUES IN USE: 7
SAPSUP: 00400001 SAPENA: 00000000
INTERFACE PACKET DROPS: 0
CONNECTION MODE: LAYER 2
IPV4 ATTRIBUTES:
VLAN ID: 663 VMAC ACTIVE: YES
VMAC ADDR: 42006A530F12 VMAC ORIGIN: OSA VMAC ROUTER: ALL
REGISTERED ADDRESSES:
IPV4 UNICAST ADDRESSES FOR ARP OFFLOAD:
ADDR: 16.11.16.105
ADDR: 16.11.17.105
ADDR: 16.11.19.105
TOTAL NUMBER OF IPV4 ADDRESSES: 3
```

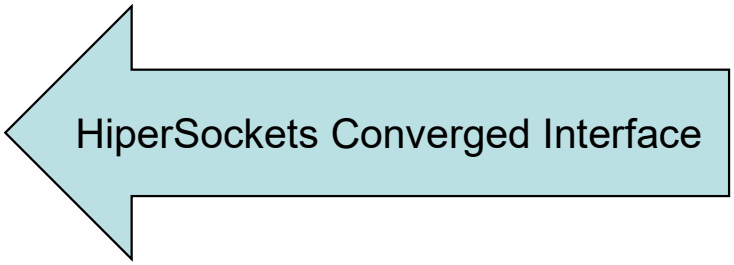
EQENET Netstat DevLinks

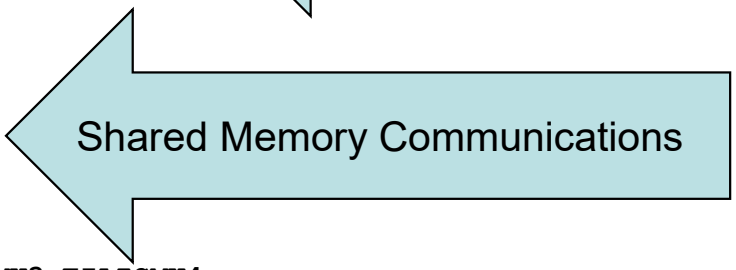
```
INTFNAME: O4ETHA1 INTFTYPE: EQENET INTFSTATUS: READY
DEVNUM: 8102 ACTDEVNUM: 8102 DEVSTATUS: READY
CHPIDTYPE: OSH CHPID: 71 PCHID: 0139
SMCD: YES SMCR: YES PNETID: NETWORK7CD
ADAPTER GEN: NETWORK EXPRESS V1.0
TRLE: IUTE8102 CODE LEVEL: 30310000000001440
PORTNAME: EZAP8102
SPEED: 0000010000 (10G)
IPBROADCASTCAPABILITY: NO
VMACADDR: 42000A69C925 VMACORIGIN: OSA VMACROUTER: ALL
CFGMTU: 9000 ACTMTU: 9000
IPADDR: 16.11.37.160/20
VLANID: 602 VLANPRIORITY: DISABLED
TOTAL READ STORAGE: 19.0M
CHECKSUMOFFLOAD: YES SEGMENTATIONOFFLOAD: YES
SECCLASS: 255 MONSYSPLEX: YES
ISOLATE: NO
MULTICAST SPECIFIC:
MULTICAST CAPABILITY: YES
GROUP REFCNT SRCFLTMD
----- ------ --------
224.0.0.5 0000000001 EXCLUDE
SRCADDR: NONE
224.0.0.1 0000000001 EXCLUDE
SRCADDR: NONE
```

Interface type = EQENET
Configured Device Number = DEVNUM
Actual Device Number = ACTDEVNUM
CHPID Type = OSH
OSA Adapter Generation = NETWORK EXPRESS V1.0
TRLE and PORTNAME dynamically created
OSA Firmware displayed = CODE LEVEL
ReadStorage dynamically managed

EQENET Netstat DevLinks (cont.)

```
INTERFACE STATISTICS:
BYTESIN = 8042214
INBOUND PACKETS = 81593
INBOUND PACKETS IN ERROR = 0
INBOUND PACKETS DISCARDED = 0
INBOUND PACKETS WITH NO PROTOCOL = 0
BYTESOUT = 1025416
OUTBOUND PACKETS = 10502
OUTBOUND PACKETS IN ERROR = 0
OUTBOUND PACKETS DISCARDED = 0
ASSOCIATED IQD CONVERGED INTERFACE: EZAIQCF9 IQC STATUS: READY
BYTESIN = 19714
INBOUND PACKETS = 121
BYTESOUT = 53236
OUTBOUND PACKETS = 464
SMCR CAPABILITY: V2
ASSOCIATED MULTI-SUBNET RNIC INTERFACE: EZARIUT1314E
ROCE PFID: 314E SMCRMTU: 4096
SMCRIPADDR: 16.11.39.160
UNASSOCIATED ISM INTERFACES: EZAISMU1 EZAISMU2 EZAISMU3 EZAISMU4
IPV4 LAN GROUP SUMMARY
LANGROUP: 00002
NAME STATUS ARPOWNER VIPAOWNER
---- ------ -------- ---------
O4ETHA1 ACTIVE O4ETHA1 NO
O4ETHA0 ACTIVE O4ETHA0 YES xxx
```





D NET,TRL,TRLE

```
IST075I NAME = IUTE3B20, TYPE = TRLE
IST1954I TRL MAJOR NODE = ISTTRL
IST486I STATUS= ACTIV, DESIRED STATE= ACTIV
IST087I TYPE = *NA* , CONTROL = MPC , HPDT = *NA*
IST1715I MPCLEVEL = EQDIO MPCUSAGE = EXCLUSIVE
IST2337I CHPID TYPE = OSH CHPID = D5 PNETID = PLEX1
IST1221I EQDIO DEV = 3B20 STATUS = ACTIVE STATE = ONLINE
IST1717I ULPID = TCPSVT ULP INTERFACE = O7ETHD0
IST2309I ACCELERATED ROUTING ENABLED
IST924I -----------------------------------------------
IST2468I INBOUND TRANSMISSION INFORMATION:
IST924I -----------------------------------------------
IST2469I QUEUE/ QUEUE STORAGE QUEUE
IST2470I ID TYPE CUR MIN MAX STATUS
IST2205I ------- -------- ------------------ ----------
IST2471I CTRL/0 CONTROL 1.0M 1.0M 1.0M ACTIVE
IST2472I READ/6 PRIMARY 8.0M 8.0M 8.0M ACTIVE
IST2472I READ/7 BULKDATA 8.0M 8.0M 8.0M ACTIVE
IST2472I READ/8 SYSDIST *NA* *NA* *NA* NOT IN USE
IST2472I READ/9 EE 4.0M 4.0M 4.0M ACTIVE
IST2472I READ/11 ZCX *NA* *NA* *NA* NOT IN USE
IST2472I READ/12 IPROUTER 4.0M 4.0M 4.0M ACTIVE
IST924I -----------------------------------------------
IST2480I CACHED READ STORAGE = 3.0M
IST2481I TOTAL READ STORAGE = 28.0M
IST924I -----------------------------------
IST2473I OUTBOUND TRANSMISSION INFORMATION:
IST924I -----------------------------------
IST2474I QUEUE/ QUEUE UNITS OF WORK QUEUE
IST2475I ID TYPE CUR AVG MAX STATUS
IST2205I ------- --------- ----------------
IST2476I CTRL/1 CONTROL 0 1 4 UNCONGESTED
IST2477I WRT/2 PRIORITY1 0 2 2 UNCONGESTED
IST2477I WRT/3 PRIORITY2 0 2 3 UNCONGESTED
IST2477I WRT/4 PRIORITY3 0 0 0 UNCONGESTED
IST2477I WRT/5 PRIORITY4 0 2 3 UNCONGESTED
```

Limitations/Guidelines

Limitations/Guidelines

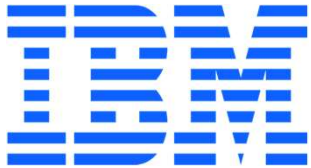
- No EQDIO support on z/OS 2.4 or a VSE client
- SNMP support for OSH will not be available at z17 GA (targeted for Sept 2025)
- OSAENTA (NetworkTraffic Analyzer) trace support will not be available
  - Alternatively, a sniffer (wireshark) trace may be used
- QDIOSYNC support will not be available
  - Automated mechanism to collect Software and Hardware traces concurrently
- NETH PFIDs (RoCE) Guideline
  - For clients using IBM Network Express 10G or 25G with z/OS Communications Server V2.5 or V3.1, it is recommended to set up a maximum of 16 NETH FIDs.
- z/OS guest deployed in zVM vSwitch environment must be Layer 3 QDIO (IPAQENET OSD CHPID OSA-Express)
  - A z/VM vSwitch attached to a Network Express card must operate in Layer-2 mode.
  - A z/OS guest cannot be deployed on this vSwitch, as the z/OS guest must use a QDIO interface (a current z/VM limitation) and z/OS does not support Layer-2 with QDIO.
  - Linux supports both modes today, so Linux guests using QDIO can use this configuration

APARs

APARs

- z/OS Communications Server V2.5 and V3.1) APARs:
  - OA64896 for SNA
  - PH54596 for TCP/IP
- IOS APAR:
  - OA63265

Thank you



Appendix – Backup Charts

SMC-R and SMC-D Definitions

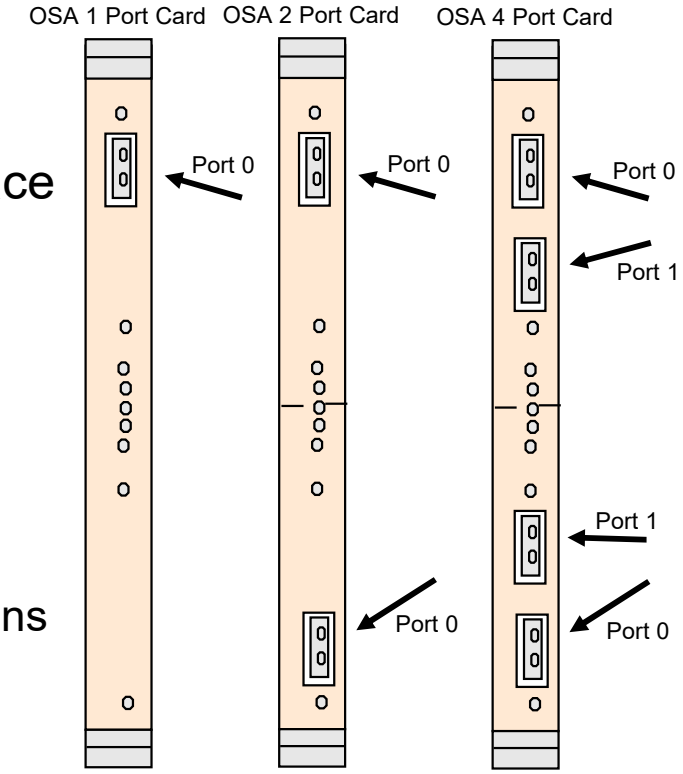
- SMC-R with RoCE
  - HCD (Hardware Configuration Definition)
    - PCHID (Physical Channel ID) – 3 digit hex value of physical slot location
    - PFID (PCIe Function ID) – unique 3 digit hex value per PCHID
    - VF (Virtual Function) – unique 2 digit decimal value per PCHID
    - Different PFID and VF per TCP/IP stack/port
    - PNET (Physical Network)(PNET on RoCE and OSA must match)
    - Port Number - **SMCv2 port is defined in HCD** and SMCv1 port is defined in PROFILE.TCPIP
  - PROFILE.TCPIP PORT 4791 UDP
    - Required for **SMCv2** traffic to be routed. Open port on Routers.
  - PROFILE.TCPIP GLOBALCONFIG SMCR
    - PFID MTU (PCIe Function ID)(Maximum Transmission Unit) – **SMCv1**
  - PROFILE.TCPIP GLOBALCONFIG SMCEID/ENDSMCEID
    - UEIDs (User-defined Enterprise IDs) – **SMCv2**
- SMC-D with ISM
  - HCD (Hardware Configuration Definition)
    - VCHID (Virtual Channel ID)
    - PNET – **required for SMCv1 but not recommended for SMCv2**
  - PROFILE.TCPIP GLOBALCONFIG SMCD
    - SYSTEMEID – causes System EID (SEID) to be generated – **SMC-Dv2 requires SEID or UEID**
  - PROFILE.TCPIP GLOBALCONFIG SMCEID/ENDSMCEID
    - UEIDs (User-defined Enterprise IDs) – **SMC-Dv2 requires SEID or UEID**

Other Optional PROFILE.TCPIP Definitions

- SMC-R with RoCE and SMC-D with ISM
  - PROFILE.TCPIP GLOBALCONFIG SMCD and SMCR
    - FIXEDMEMORY - Specifies the maximum amount of 64-bit storage that the stack can use for the send and receive buffers that are required for SMC-D and SMC-R communications. 256 megabytes default.
    - TCPKEEPMININTERVAL - This interval specifies the minimum interval that TCP keepalive packets are sent on the TCP path of an SMC-D or SMC-R link. 300 seconds default.
  - PROFILE.TCPIP GLOBALCONFIG
    - AUTOCACHE - Specifies whether this stack caches unsuccessful attempts to use SMC communication. AUTOCACHE is the default.
    - AUTOSMC - Specifies whether this stack monitors inbound TCP connections to dynamically determine whether SMC is beneficial for a local TCP server application. AUTOSMC is the default.
    - SMCPERMIT/ENDSMCPERMIT - Specifies the SMC filter that allows SMC negotiation with peers within the listed TCP/IP address(es)/subnet(s).
    - SMCEXCLUDE/ENDSMCEXCLUDE - Specifies the SMC filter that prevents SMC negotiation with peers within the listed TCP/IP address(es)/subnet(s).

OSD Customization Requirements

- QDIO OSA Customization (TCP/IP only)
  - HCD (IOCP) CHPID type OSD and 3 devices
    - 1 Control (Data Path) device per TCP/IP Interface
    - 1 Read device per LPAR
    - 1 Write device per LPAR
  - VTAM Customization
    - Define TRL
  - TCP/IP Customization
    - TCP/IP Profile
      - Define INTERFACE
      - Configure BeginRoutes or OMPROUTE definitions
      - Define START statement



•Prior to OSA-Express3 each OSA provides a single port per CHPID.
•OSA four port cards provide two ports per CHPID.
•No difference between HCD for OSA with single port per CHPID and two ports per CHPID.

OSA HCD

- QDIO OSA HCD (IOCP)
  - Channel path type=OSD
  - CNTLUNIT UNIT=OSA
  - Device type=OSA
    - Minimum of 3 Devices per z/OS
      - 1 Device for Read processing
      - 1 Device for Write processing
      - 1 Device for the Data Path per INTERFACE
      - i.e.. z/OS with CINET environment with 2 TCP/IP stacks would require 4 devices. (1 Read + 1 Write + 2 Data Path = 4)

HCD (IOCP) parameter CHPARM=00 enables Priority Queuing.

OSA mode OSM does not support Priority Queuing so it is automatically set to CHPARM=02 for OSM.

CHPARM=00 for OSA mode OSC is used for TN3270E protocol.
CHPARM=40 for OSA mode OSC is used for 3215 protocol to TPF.

Example of an OSA-Express2 CHPID or two ports on one CHPID of an OSA-Express3:

```
CHPID PATH=(CSS(0.1),02),SHARED, PARTITION=((CSS(1),(A12),(=))),
    PCHPID=1C0,TYPE=OSD
CNTLUNIT CUNUMBR=2980,PATH=((CSS(0),02),(CSS(1),02)),UNIT=OSA
IODEVICE ADDRESS=(2980,015),UNITADD=00,CUNUMBR=(2980),UNIT=OSA
IODEVICE ADDRESS=298F,UNITADD=FE,CUNUMBR=(2980),UNIT=OSAD
```

Another example of two ports on one CHPID of an OSA-Express3:

```
CHPID PATH=(CSS(0.1),02),SHARED, PARTITION=((CSS(1),(A12),(=))),
    PCHPID=1C0,TYPE=OSD
CNTLUNIT CUNUMBR=3980,PATH=((CSS(0),02),(CSS(1),02)),UNIT=OSA
IODEVICE ADDRESS=(3980,015),UNITADD=00,CUNUMBR=(3980),UNIT=OSA
IODEVICE ADDRESS=398F,UNITADD=FE,CUNUMBR=(3980),UNIT=OSAD
IODEVICE ADDRESS=(4980,015),UNITADD=20,CUNUMBR=(3980),UNIT=OSA
```

VTAM TRL

- QDIO OSA VTAM TRL major node member

```
TRL VBUILD TYPE=TRL
   trl_name TRLE LNCTL=MPC,
     READ=(xxx),
     MPCLEVEL=QDIO,
     WRITE=(yyy),
     DATAPATH=(zzz),
     PORTNAME=device_name,
     PORTNUM=1
```

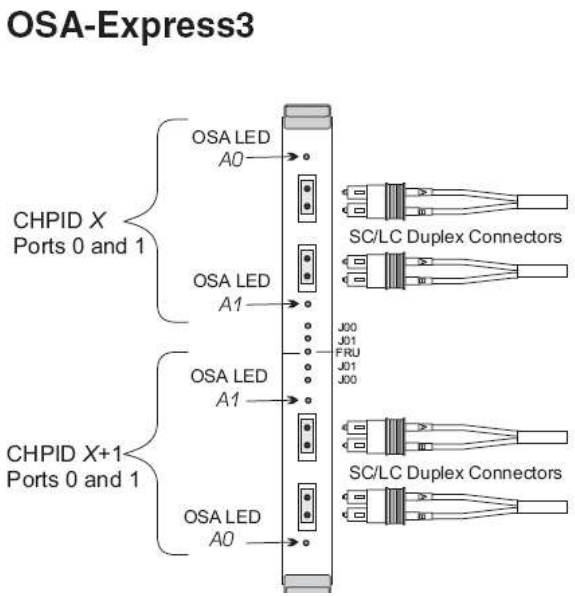
PORTNUM is always 0 for single port per CHPID OSA.
–When PORTNUM=1 is defined for single port per CHPID OSA it is ignored.
On OSA two-port/CHPID cards code PORTNUM=0 for port 0 and PORTNUM=1 for port 1.

All z/OS systems that share a port must define the port with the same PORTNAME.
PORTNAME relief for z/VM and zLinux APAR PQ73878

Note: z/OS PORTNAME must be unique. There is a subtle difference in PORTNAME support between OSA-E2 and OSA-E3. Two OSA ports may use the same PORTNAME if they are on the same OSA card but on different CHPIDs and not both defined to the same VTAM. With OSA-E2 the ports are always on different CHPIDs but with OSA-E3 two ports could be on the same CHPID.

OSA TCP/IP Interface

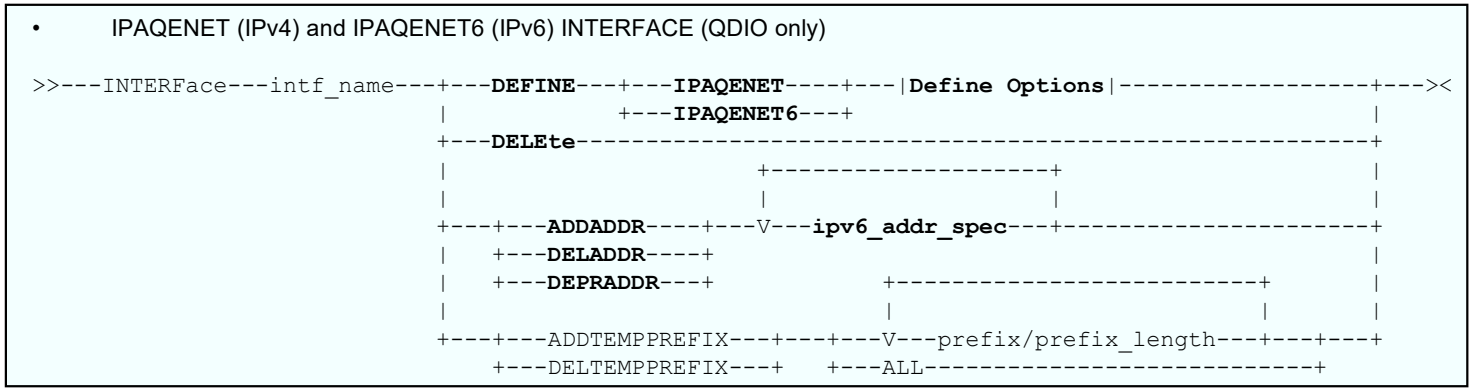
- QDIO TCP/IP Profile INTERFACE (IPv4)
  INTERFace intf_name intf_action IPAQENET…
  START interface_name
- QDIO TCP/IP Profile INTERFACE (IPv6)
  INTERFace intf_name intf_action IPAQENET6…
  START interface_name
- Port name and device name must match:
  - QDIO
    - TRLE port name = INTERFACE port name
- TRLE port name must match in all z/OS TCP/IP stacks for a shared OSA port.
- Per TCP/IP Stack:
  - Only one TRLE per port.
  - Only one INTERFACE per TRLE without VMAC.
- QDIO MPCIPA Link Types
  - IPAQENET/6
    - Ethernet QDIO OSA
    - QDIO always uses ETHEROR802.3.



- If the OSA is configured for both IPv4 and IPv6 for a stack, then the stack must use one VMAC on the INTERFACE statement for IPv4 usage, and a different VMAC on the INTERFACE statement for IPv6 usage.

Interface Action Types

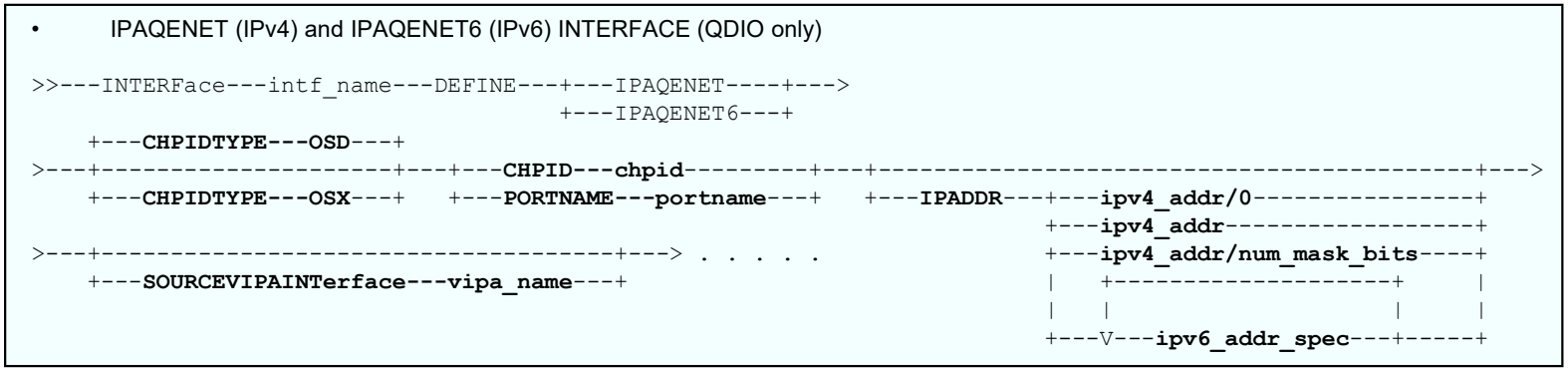
- IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)



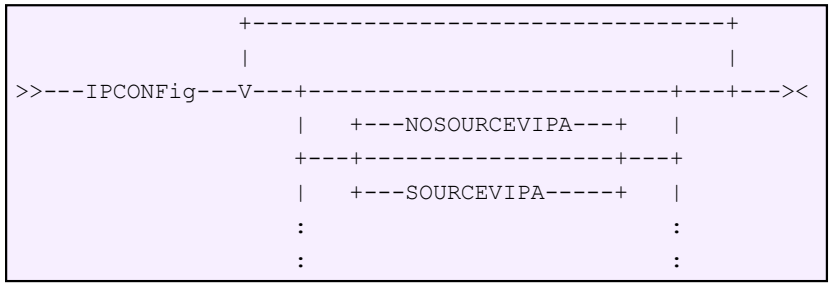
- IPAQENET (in z/OS V1.10+) and IPAQENET6 intf_action types:
  - DEFINE
    - Adds the Interface to the list of defined adapters.
  - DELEte
    - Removes the Interface from the list of defined adapters.
    - DELETE does not have any parameters associated with it.
- Additional IPAQENET6 only intf_action types:
  - ADDADDR
    - Adds the address to the defined Interface definition.
  - DELADDR
    - Removes the address from the defined Interface definition.
  - DEPRADDR
    - Deprecates the address in the Interface definition. This makes the address less preferred. See the “Default address selection” section of the “IPv6 Network and Design Guide, SC31-8885”.

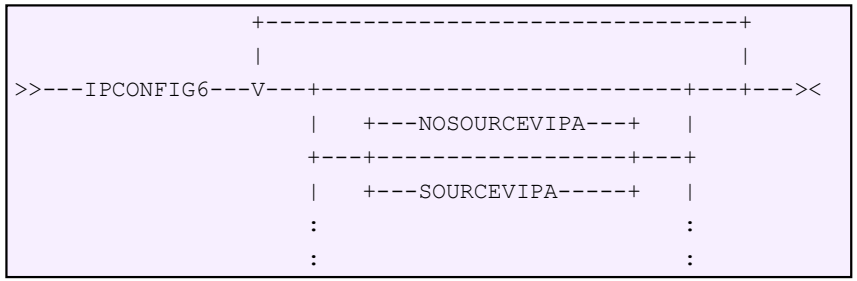
CHPID, IP Address, Port Name, and Source VIPA

- IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)



- CHPIDTYPE
  - Indicates either OSD or OSX mode.
- CHPID chpid
  - Identifies the CHPID for the interface. A 2-character hexadecimal value (00 - FF).
- ipaddr_spec
  - Specifies the ipv4_addr, ipv4_addr/mask, ipv6_addr, or prefix/prefix_length.
- PORTNAME port_name
  - Port name and device name must match between TRLE, DEVICE, and INTERFACE.
  - Interface and Portname must be different.





- SOURCEVIPAINTerface vipa_name
  - Specifies which static VIPA interface is to be used for SOURCEVIPA.
  - Requires IPCONFIG or IPCONFIG6 SOURCEVIPA.

Traffic Path In/Out of z/OS

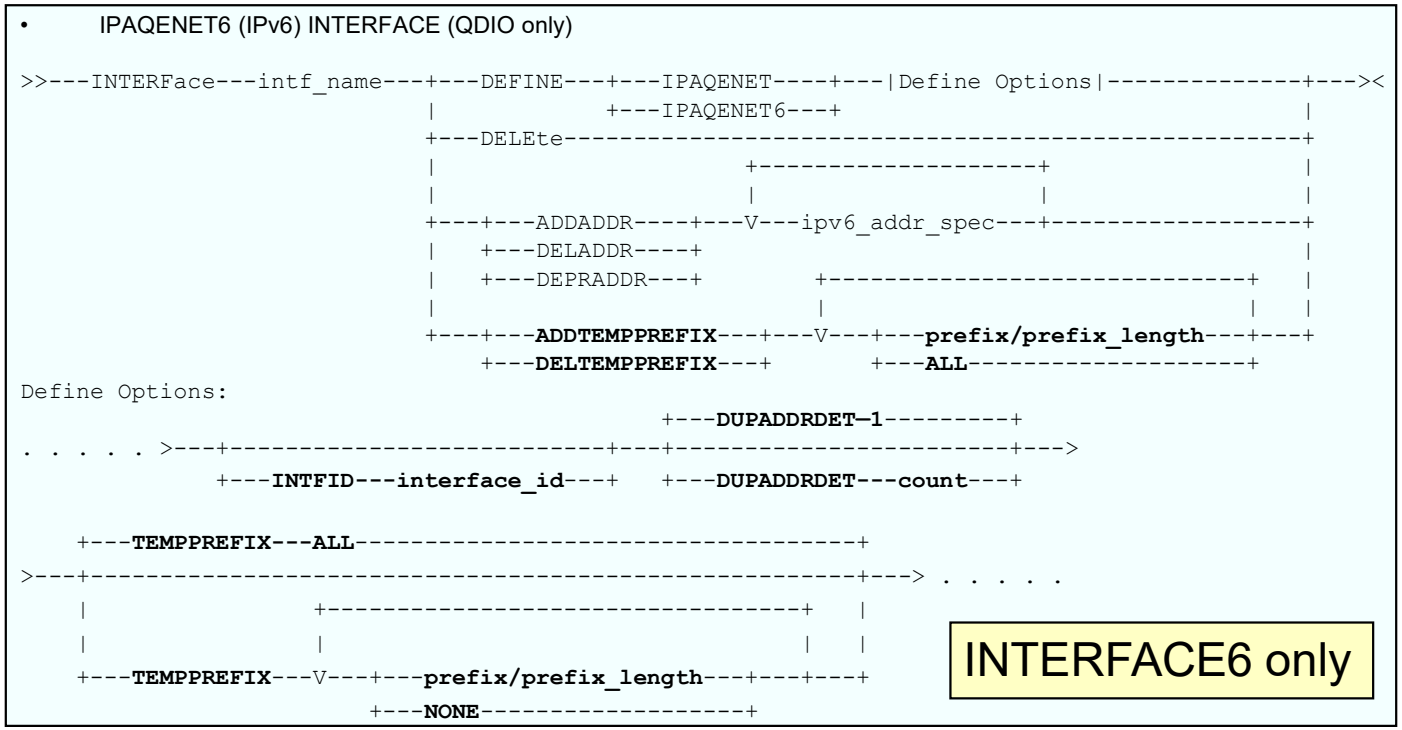
- Inbound Connections
  - When a connection is initiated from a remote node, the remote node sends a connection packet to a z/OS application.
    - Typically the z/OS TCP application swaps the source and destination IP address from the received packet to be used in the response packet.
      - Takes the source addr from received packet and uses it as destination addr in the response packet.
      - Takes the destination addr from the received packet and uses it as source addr in the response packet.
- Outbound Connections
  - When a connection is initiated from a z/OS application the source IP address is determined by the source IP address selection algorithm.
    - See “Source IP Address Selection” section in the IP Configuration Guide manual.
  - When a connection is initiated from a z/OS application the destination IP address is either passed to the application in the connection command (ie. ftp 9.15.42.10) or is determined by domain name resolution of host name (ie. ftp wscftpsrv).
- Routing Table is used to send packet
  - Destination IP address is used with Routing Table (may contain a combination of static and dynamically learned routes), to determine which network path to send packet over.

Source IP Address Selection

- As per the IP Configuration Guide…
- TCP/IP determines the source IP address for a TCP outbound connection, or for a UDP or RAW outbound packet, using the following sequence, 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. PORT profile statement with the BIND parameter
  5. SRCIP profile statement (TCP connections only)
  6. TCPSTACKSOURCEVIPA parameter on the IPCONFIG or IPCONFIG6 profile statement (TCP connections only)
  7. SOURCEVIPA: static VIPA address from the HOME list or from the SOURCEVIPAINTERFACE parameter
  8. HOME IP address of the link 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.

Duplicate Address Detection Count, Interface ID, and Temporary Address Prefix

- IPAQENET6 (IPv6) INTERFACE (QDIO only)

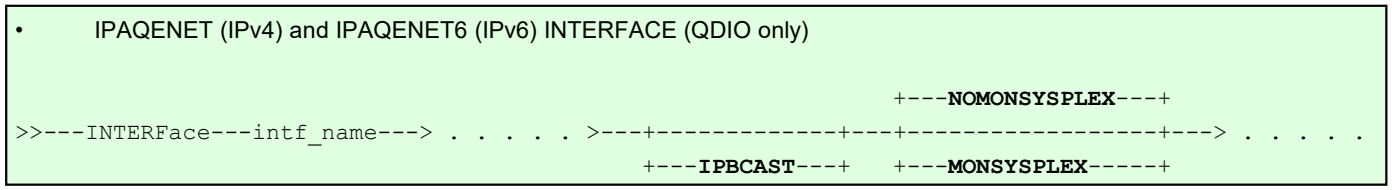


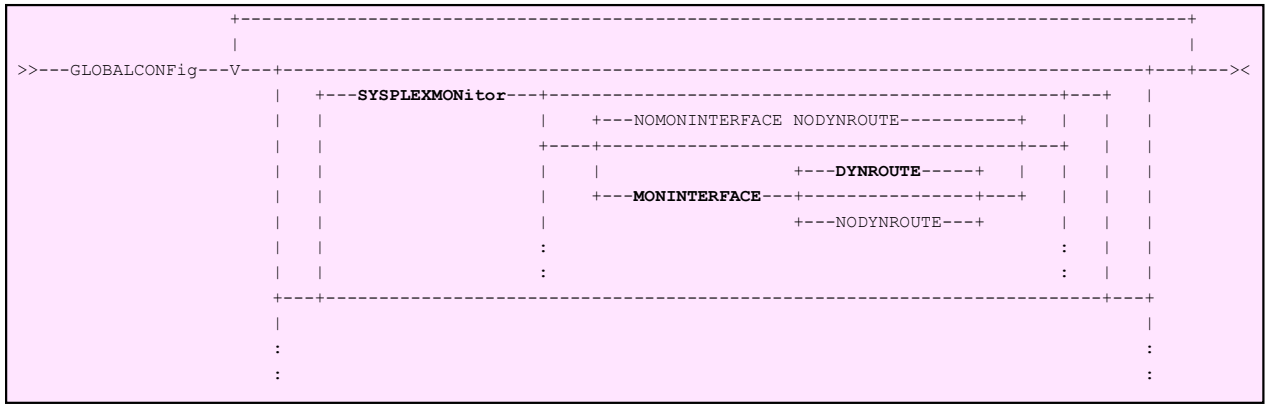
- DUPADDRDET count_num
  - Number of times duplicate address detection is done.
- INTFID interface_id
  - 64-bit interface identifier in colon-hexadecimal format.
  - Interface ID is either manually defined by this parameter or TCP/IP builds the Interface ID using information from the OSA.
  - Interface ID is used to form the link-local address for the interface, and is also appended to any prefixes for the interface (either manually configured (by IPADDR or ADDADDR), or learned over the interface by router advertisements) to form complete IPv6 addresses.

- TEMPPREFIX ALL/NONE/prefix/prefix_length (Default is ALL)
  - ALL causes temporary addresses to be generated for all prefixes learned over this interface by router advertisements.
  - NONE causes no IPv6 temporary addresses to be generated.
  - prefix/prefix_length specifies the set of prefixes for which temporary IPv6 addresses will be generated.
  - Temporary addresses are generated only on an interface that is enabled for stateless address autoconfiguration.
    - Stateless address autoconfiguration is enabled for an interface if no address or prefix is specified with the IPADDR keyword.
  - Temporary addresses are generated only when the TEMPADDRS keyword is specified on the IPCONFIG6 statement.
  - The job name of an application must be in the SRCIP statement block with a value of TEMPADDRS to cause a temporary IPv6 address to be preferred over a public IPv6 address as the source IP address for the application; otherwise, the default source address selection algorithm prefers public IPv6 addresses over temporary addresses.
  - Temporary prefixes may be added using ADDTEMPPREFIX and deleted using DELTEMPPREFIX.

Broadcasts, and Sysplex Monitor

- IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)





- IPBCAST
  - Enables IP broadcasts over this link. Without IPBCAST no IP broadcast will be passed over this link.
- MONSYSPLEX
  - Specifies whether or not sysplex autonomics should monitor the link's status.
    - MONINTERFACE is required on GLOBALCONFIG SYSPLEXMONITOR statement.
  - Dynamic routes over this link may be monitored.
    - Requires MONSYSPLEX and DYNROUTE on the GLOBALCONFIG SYSPLEXMONITOR statement.
  - NOMONSYSPLEX is the default.
  - See VIPA presentation for more information about Sysplex Autonomics:
    - http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS789

Security Class Parameter

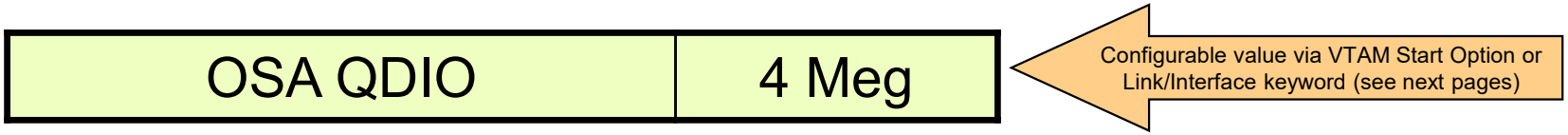
```
•       IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)

                                          +---SECCLASS---255--------------+
>>---INTERFace---intf_name---> . . . . . >---+-------------------------------+---> . . . . .
                                          +---SECCLASS---security_class---+
```

```
                +-----------------------------------------------------------+
                |                                                           |
>>---IPCONFIG6---V---+-----------------------------------------------------+---+--->< 
        +----+-------------------------------------------------------+---+
        |    |               +---OSMSECCLASS 255----------------+   |   |
        |    +---IPSECURITY---+-----------------------------------+---+   |
        |                    +---OSMSECCLASS---security_class---+       |
        :                                                               :
        :                                                               :
```

- SECCLASS security_class
  - Used for Multi-Level Security.
  - Security class for IP filtering with this interface.
  - The matching policy action is applied when the SECCLASS parameter matches the SecurityClass parameter defined on the policy IPsec condition IpService statement.
  - TCP/IP stack ignores this value if IPSECURITY is not specified on the IPCONFIG or IPCONFIG6 statement.

OSA QDIO Read Storage

- Amount of storage for read processing:



- The storage used for read processing is allocated from the CSM data space 4K pool, and is fixed storage backed by 64-bit real. (CSM fixed storage defined in PARMLIB member IVTPRMxx)
- OSA QDIO
  - 64 SBALs (storage block address lists) x 64K = 4M

VTAM Start Options to Define OSA Storage

- OSA QDIO Read Storage VTAM Start Option QDIOSTG
  - Defines how much storage VTAM keeps available for read processing for all OSA QDIO devices

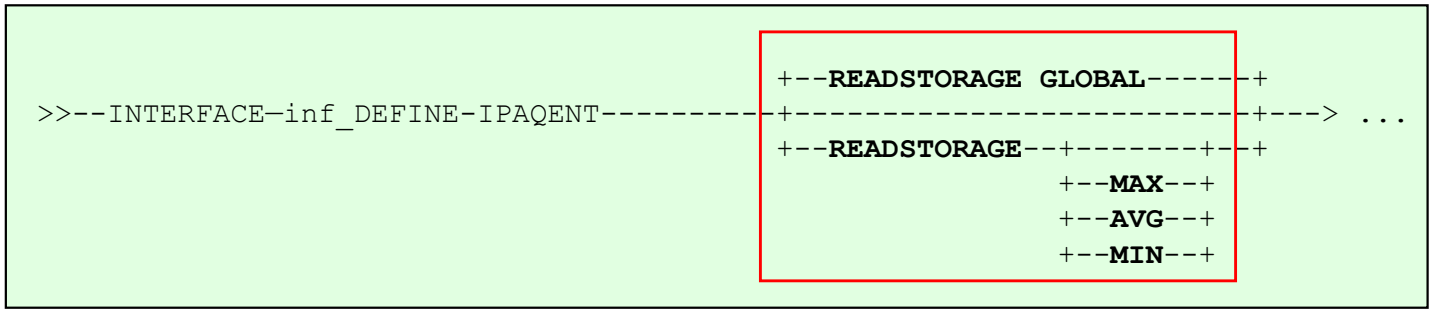
```
        +--QDIOSTG=-MAX---------+
>>------+-----------------------+--->
        +--QDIOSTG=--+--MAX--+--+
                     +--AVG--+
                     +--MIN--+
                     +--nnn--+
```

MAX	64 SBALs x 64K = 4M
AVG	32 SBALs x 64K = 2M
MIN	16 SBALs x 64K = 1M

Default

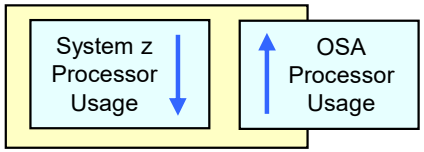
- Storage units are defined in terms of QDIO SBALs (QDIO read buffers)
  - nnn is the exact number of SBALs in the range 8-126
  - MAX allows for the best performance (for example, throughput), but requires more storage.
  - MIN may be used for devices with lighter workloads or where system storage might be constrained.
  - The amount of storage used is times the number of active QDIO data devices.
- Start Option defaults are appropriate for most environments
  - Review CSM specifications in PARMLIB member IVTPRMxx and increase, if appropriate
  - Use the D NET,CSM to display CSM usage
  - Modify storage settings using Start Options, as appropriate
  - Use VTAM tuning stats to evaluate needs and usage. Under a typical workload, the NOREADS counter should remain low (close to 0). If this count does not remain low you may need to consider a higher setting for QDIOSTG.

Read Storage Parameter



- READSTORAGE
  - Defines the amount fixed storage for read processing.
  - READSTORAGE must match between LINK and INTERFACE for the same OSA.
- Overrides VTAM Start option QDIOSTG for a specific QDIO device.
- Global causes the QDIOSTG VTAM start option values to be used.
  - This is the default.
- MAX, AVG, and MIN
  - Causes the MAX, AVG, or MIN VTAM Start option MAX, AVG, or MIN values to be used.

ARP, Checksum, and Segmentation Offload

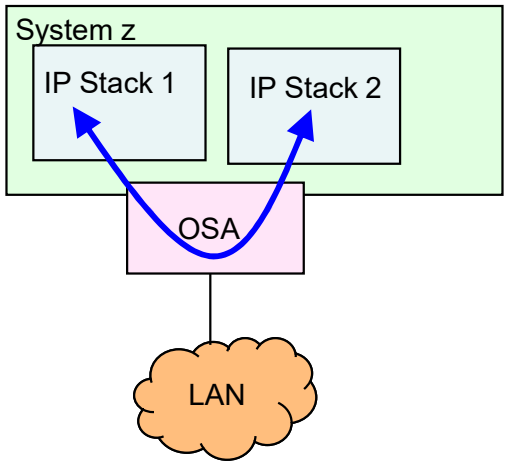



```
                +-----------------------------------------+
                |                                         |
>>---IPCONFig---V---+-----------------------------------+---+---><
                    |    +---CHECKSUMOFFLoad-----+       |
                    +----+----------------------+-------+
                    |    +---NOCHECKSUMOFFLoad---+       |
                    |    +---NOSEGMENTATIONOFFLoad---+   |
                    +----+---------------------------+---+
                    |    +---SEGMENTATIONOFFLoad-----+   |
                    :                                    :
                    :                                    :
```

```
                +-----------------------------------------+
                |                                         |
>>---IPCONFIG6---V---+-----------------------------------+---+---><
                     |   +---CHECKSUMOFFLoad-----+        |
                     +---+----------------------+--------+
                     |   +---NOCHECKSUMOFFLoad---+        |
                     |    +---NOSEGMENTATIONOFFLoad---+   |
                     +----+---------------------------+---+
                     |    +---SEGMENTATIONOFFLoad-----+   |
                     :                                    :
                     :                                    :
```

- The following OSA Offload Support avoids CPU processing.
- IPv4 ARP Offload
  - OSA automatically offloads ARP processing.
  - Use the NETSTAT command to display the OSA ARP cache.
  - Use the vary command to flush the ARP cache on the OSA.
    - **Vary TCPIP,proc_name,PURGECache,intf_name**
- Checksum Offload
  - A checksum of the packet data is calculated and sent with the packet to provide integrity of the data.
  - OSA Checksum Offload processing is controlled by the IPConfig and IPConfig6 statements.
- Segmentation Offload
  - Also known as Large Send for TCP/IP traffic, Segmentation Offload allows larger amounts of data to be sent by the TCP/IP stack because the OSA provides the segmentation of that data.
  - OSA Segmentation Offload processing is controlled by the IPConfig and IPConfig6 statements.
  - Always check for the latest PSP bucket and OSA driver levels.

Routing for Shared OSA

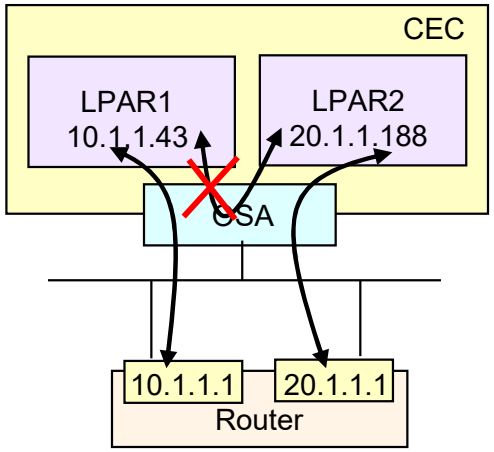


- All IP addresses in HOME list are added to OSA Address Table (OAT)
- When a packet is sent from one of the systems sharing the OSA and the destination is an IP address in the OAT, the packet is sent directly to the destination without going out onto the LAN.

Disable LPAR to LPAR Traffic via Shared OSA

```
•       IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)

                                                          +---NOISOLATE---+
>>---INTERFace---intf_name---DEFINE---+---IPAQENET----+---+---------------+---> . . . . .
                                      +---IPAQENET6---+   +---ISOLATE-----+
```

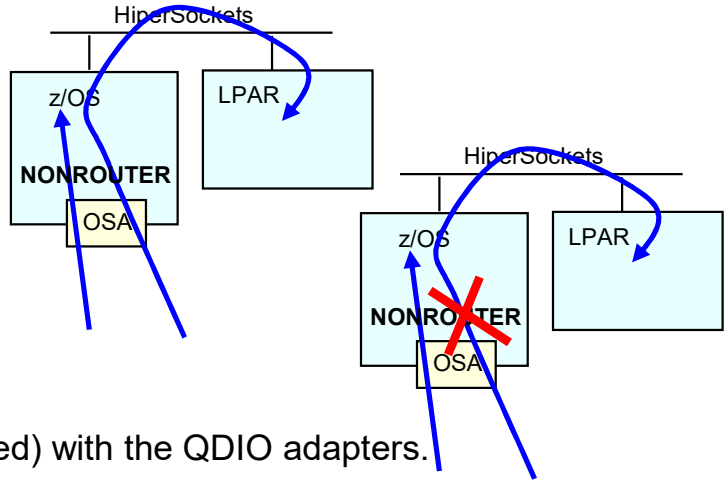
- Isolate Option
  - ISOLATE/NOISOLATE option on QDIO network interface definition.
  - Only OSA local routing, without flowing out onto the LAN, is disabled.
  - LPAR to LPAR traffic may still flow over the OSA if it is sent out onto the LAN to a router and then back in over the same OSA.
- OSA local routing can in some scenarios be seen as a security exposure.



Be careful using ISOLATE if using OSPF and share a subnet between TCP/IP stacks that share the OSA.

PRIRouter, SECRouter, NONRouter

```
•       IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)

                                                          +---NONRouter---+
>>---INTERFace---intf_name---DEFINE---+---IPAQENET-----+---+---------------+---> . . . . .
                                      +---IPAQENET6----+   +---PRIRouter---+
                                                          +---SECRouter---+
```



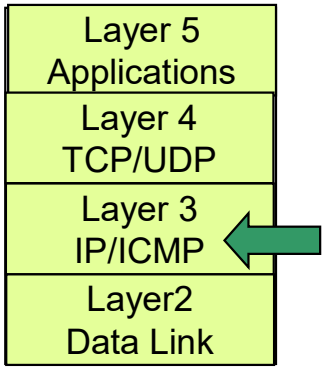
- All IP addresses in a TCP/IP HOME list are registered (dynamically downloaded) with the QDIO adapters.
  - HOME changes are automatically sent to QDIO adapters.
- If the OSA receives any packets with its MAC as the destination and a destination IP address that is "unknown" (meaning not an IP address in the HOME list), then OSA does the following:
  - If PRIRouter is defined (assuming the OSA is started to that stack) then all "unknown" packets are sent to the PRIRouter stack.
  - If PRIRouter is not defined (or the OSA is not started to any stack with PRIRouter)(could be that PRIRouter is coded but that OSA connection is down due to failure or other outage) then if SECRouter is coded all "unknown" packets are sent to the SECRouter stack. If multiple SECRouters then a random (unpredictable) stack with SECRouter coded will be sent the "unknown" packets.
    - There is no way to set the order of precedence for the secondary routers.
  - If only NONRouter is defined (or any PRIRouter and SECRouter connection are down due to failure or other outage) then all "unknown" packets are discarded by the OSA. NONROUTER is the default.
- Non-QDIO OSA (OSE mode) may define PRIROUTER and SECROUTER via OSA/SF.
- IPCONFIG DATAGRAMFWD
  - PRIRouter is used when traffic is routed through the stack to another stack. Keep in mind that if one stack is used to route to other stacks IPCONFIG DATAGRAMFWD is required.
- PriRouter, SecRouter, NonRouter definition is ignored when VMAC parameter is defined.
  - **Recommendation: Use VMAC for shared OSA ports rather than PRIROUTER/SECROUTER.**

VMAC (Virtual MAC)

- IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)

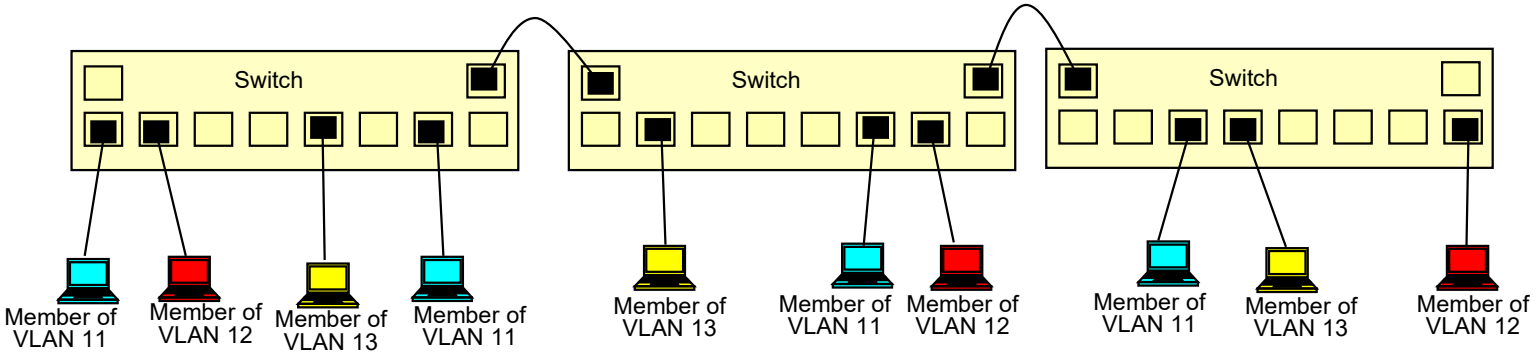
```
                                                  +---VMAC---ROUTEALL-------------------------------+
>>---INTERFace---intf_name---DEFINE---+---IPAQENET----+---+-----------------------------------------------+---> . . . . .
                                      +---IPAQENET6---+   |                               +---ROUTEALL---+   |
                                                  +---VMAC---+-------------+---+--------------+---+
                                                             +---macaddr---+   +---ROUTELCL---+
```

- VMAC mac_addr ROUTEALL/ROUTELCL
  - Indicates virtual MAC address. OSA uses this address rather than the physical MAC address for all IP packets (Layer 3) to and from this TCP/IP stack.
  - If mac_addr is not coded, then the OSA generates a virtual MAC address.
    - Unless the virtual MAC address must remain the same even after TCP/IP restart, configure VMAC without mac_addr.
  - NONROUTER, PRIROUTER, and SECROUTER are ignored for an OSA if the VMAC parameter is configured.
    - Recommendation: Use VMAC for shared OSA ports rather than PRIROUTER/SECROUTER.
  - ROUTEALL causes all IP traffic destined to the virtual MAC to be forwarded to the TCP/IP stack. The default.
  - ROUTELCL causes only the traffic destined to the virtual MAC and whose destination IP address is registered to the OSA (all active IP addresses on the TCP/IP stack), to be forwarded to the TCP/IP stack.
  - VMAC is required to define multiple VLAN IDs for IPv4 or IPv6, from a single stack for a given OSA port.



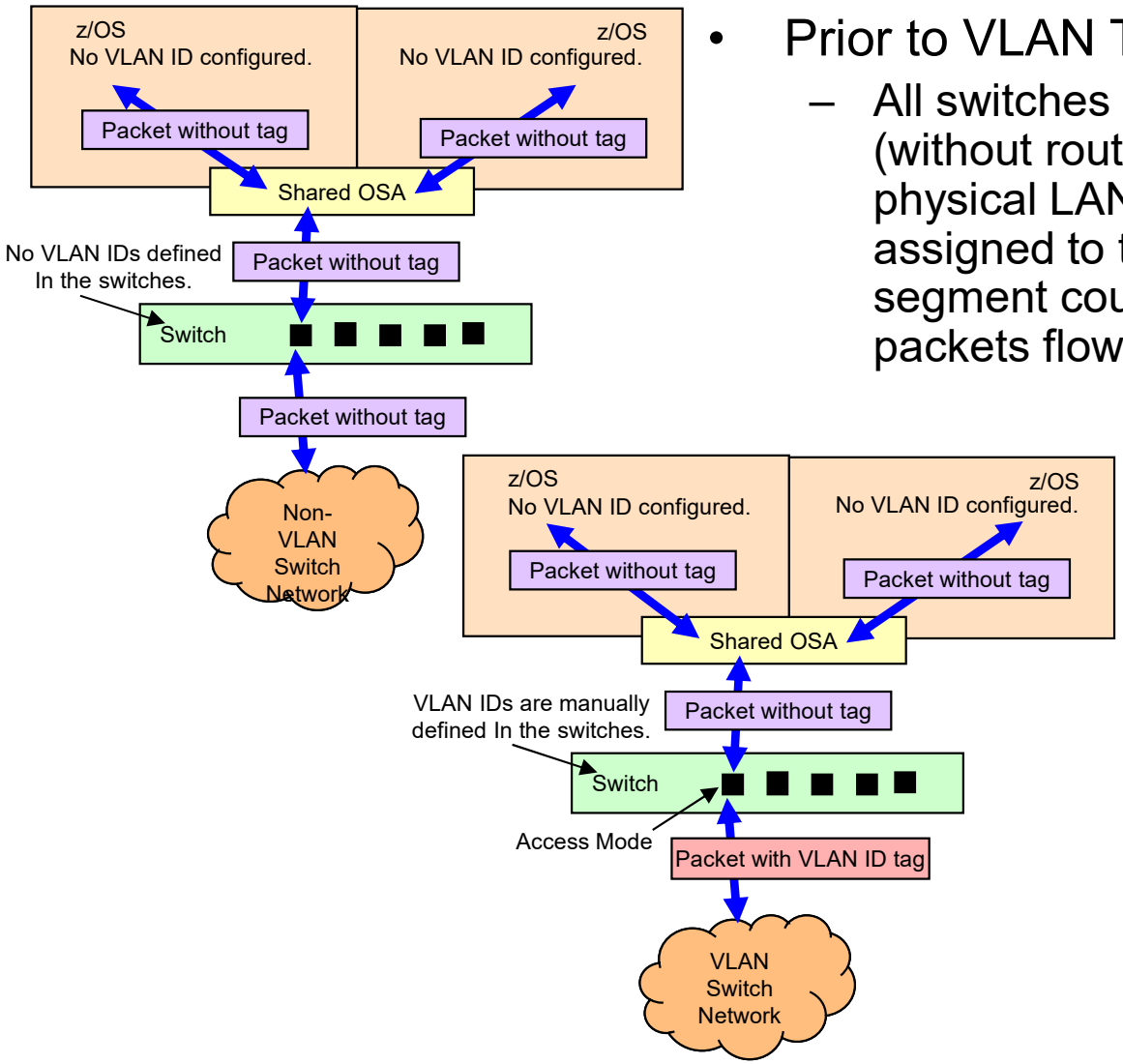
OSA Layer 2 is supported by z/VM and zLinux but is not supported by z/OS.

What is a VLAN?



- A VLAN is a switched network that is logically segmented on an organizational basis, by functions, project teams, or applications rather than on a physical or geographical basis.
- Reconfiguration of the network can be done through software rather than by physically unplugging and moving devices or wires.
- A VLAN can be thought of as a broadcast domain that exists within a defined set of switches.
- A VLAN consists of a number of end systems, either hosts or network equipment (such as bridges and routers), connected by a single bridging domain.
- VLANs are created to provide the segmentation services traditionally provided by routers in LAN configurations.
- None of the switches within the defined group will bridge any frames, not even broadcast frames, between two VLANs.
  - IP Router is needed to communicate between VLANs.

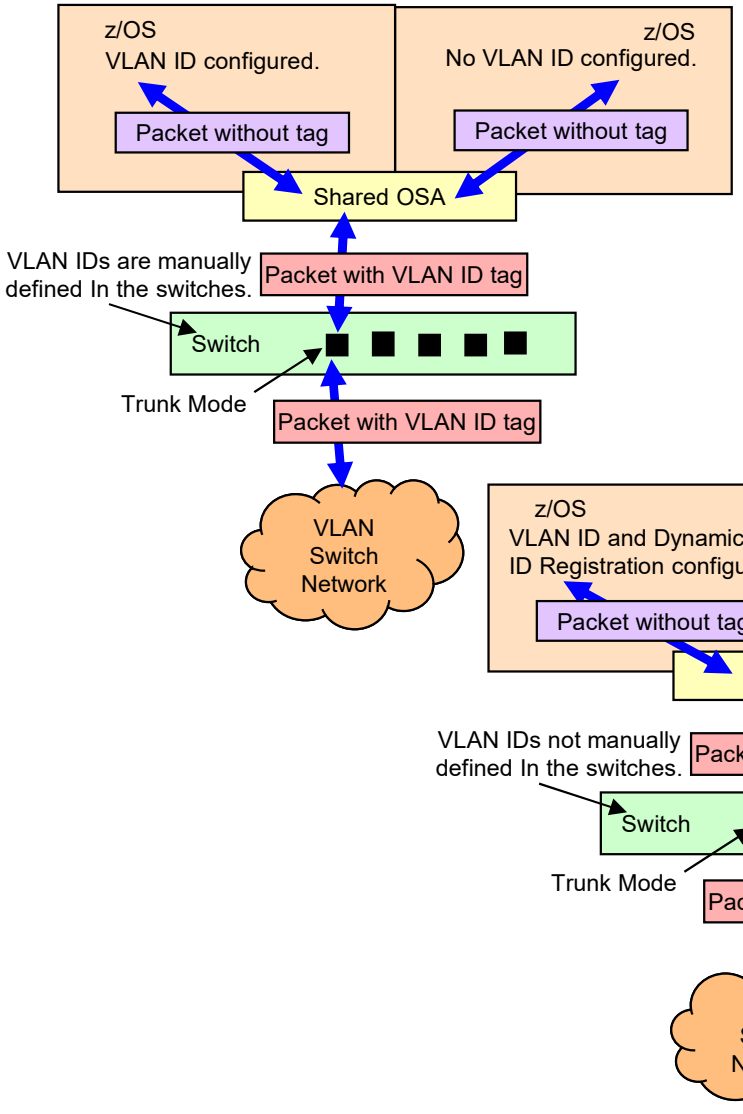
When z/OS is VLAN “un-aware”



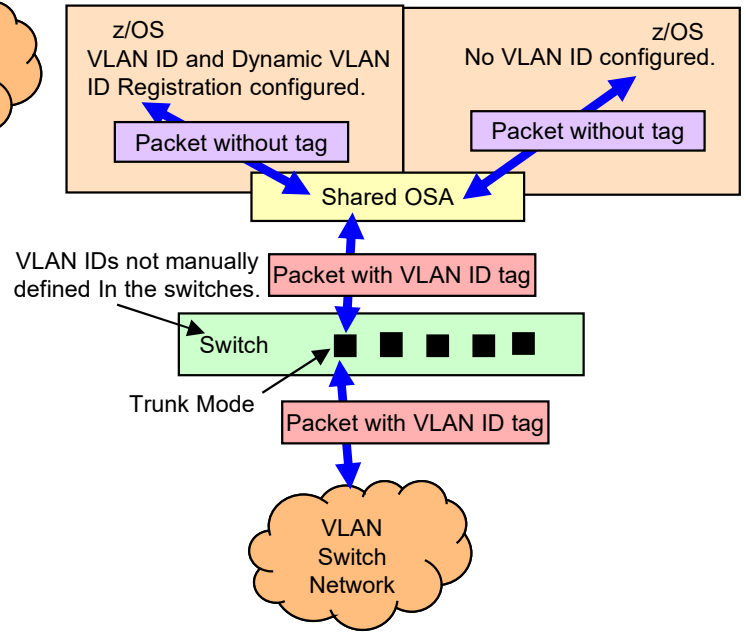
- Prior to VLAN Technology
  - All switches that were attached together (without routers in between them) formed one physical LAN segment with one IP subnet assigned to them. All devices on a LAN segment could potentially have access to all the packets flowing on the segment.

- VLANs Defined on Switches
  - z/OS may still be VLAN “un-aware”.
  - The switch port that OSA attaches to should be configured in Access Mode with a certain VLAN ID assigned.
  - The switch itself manages the VLAN ID tagging of the packets.

When z/OS is VLAN “aware”



- VLAN ID configured INTERFACE
  - The switch port that OSA attaches to must be configured in Trunk Mode.
  - OSA learns the VLAN ID from the stack and manages tagging the packets with the appropriate VLAN ID.
  - Packets from stacks that do not configure a VLAN ID (still VLAN “un-aware”) are part of the default VLAN ID (usually VLAN ID 1).
  - Multiple VLAN IDs per stack/OSA port per IP version (IPv4 or IPv6) requires z/OS V1.10 and VMACs.



- VLAN ID and Dynamic VLAN ID Registration Defined on INTERFACE
  - Rather than manually configure the supported VLAN IDs per switch port, the switch learns the VLAN IDs for the port from the OSA.

OSA QDIO VLAN Support

```
•       IPAQENET (IPv4) and IPAQENET6 (IPv6) INTERFACE (QDIO only)

                                                                +---NODYNVLANREG---+
>>---INTERFace---intf_name---> . . . . . >---+-----------------+---+-----------------+---> . . . . .
                                        +---VLANID---id---+   +---DYNVLANREG-----+
```

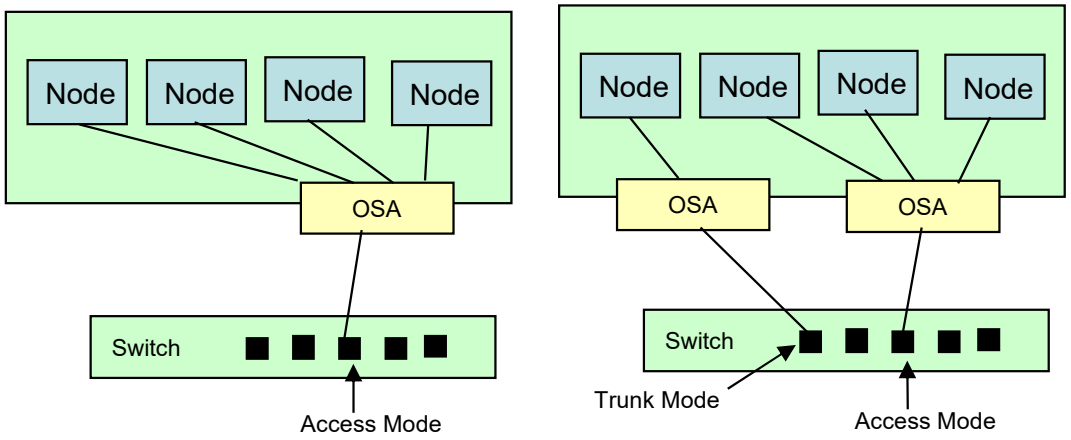
Multiple VLAN IDs per IP version per stack/OSA port (Interface Only – not supported on Link)
-Maximum of 8 VLAN IDs per IP version (IPv4 or IPv6) per OSA port per stack.
-Different VMACs are required.

- VLANID id_number
  - Specifies the VLAN ID tag for this link.
- DYNVLANREG/NODYNVLANREG
  - Dynamic registration of VLAN ID (GVRP).
    - Dynamic registration of VLAN IDs is handled by OSA and switch. Both must be at a level with the hardware support for dynamic VLAN ID registration.
  - DYNVLANREG specifies that if a VLAN ID is configured for this link, it is dynamically registered with physical switches on corresponding LAN.
    - This parameter is only applicable if a VLAN ID is specified.
  - NODYNVLANREG specifies that if VLAN ID is configured, it must be manually registered with switches on corresponding LAN. This is the default.
- **VMAC is required to define multiple VLAN IDs for IPv4 or IPv6, from a single stack for a given OSA port.**

z/OS Support of OSA VLAN IDs

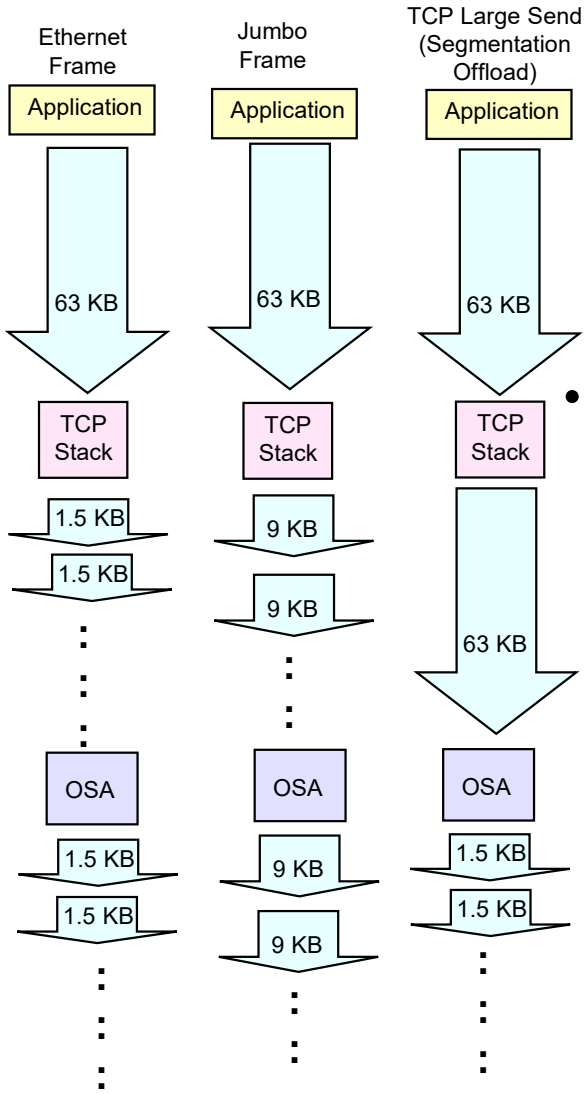
- z/OS TCP/IP supports configuring the VLAN ID to be used on OSA connections.
  - z/OS may configure the VLAN ID but it is OSA that adds/removes the VLAN ID tag to the packets.
  - Conforms to the IEEE 802.1Q standard
- A Switch may configure a port in Trunk mode or Access mode.
  - Trunk mode
    - VLAN ID is defined by the end device, either configured on z/OS or defaulted by the OSA.
    - Requires VLAN ID tagged packets.
  - Access mode
    - VLAN ID is controlled by the switch rather than the end device. Any VLAN ID configured by z/OS is ignored.
- z/OS VLAN Rules:
  1. An OSA should either be:
     - Attached to a switch port in trunk mode if any of the stacks that share the OSA have a VLAN ID configured, or
     - Attached to a switch port in access mode and each stack that shares the OSA should not have a VLAN ID configured.
  2. As with any IP network, separate VLANs should be treated like separate physical networks and have separate subnets assigned.
  3. 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.
  4. When a TCP/IP stack has access to multiple OSA ports that are on the same physical LAN, and a VLAN ID is configured on any of the OSA ports, it is recommended that this stack configure a VLAN ID for all OSA ports on the same physical LAN. Do not mix VLAN and no-VLAN on the same physical network accessed by a single stack through multiple OSA ports.
  5. When multiple INTERFACE statements are defined on a single stack for a single OSA port and a single IP version (IPv4 or IPv6), the VLAN IDs must be unique, and the INTERFACE definition will be rejected if the VLAN ID is omitted.
     - The VLAN ID, VMAC, and IP subnet values must be unique per IP version (IPv4 or IPv6) for multiple INTERFACE statements for a single OSA port defined on a single TCP/IP stack.
     - For parallel interfaces into the same IP subnet/VLAN ID from a single TCP/IP stack, multiple OSA ports are required.
  6. The requirement for a unique VLAN ID per INTERFACE statement rule only applies within a single stack. Each stack on a shared OSA port is completely independent of other stacks sharing the OSA port. Multiple stacks may define the same VLAN ID or different VLAN IDs for the same shared OSA port.

OSA VLAN Migration



- Migration of z/OS VLAN “unaware” to z/OS VLAN “aware”
- Switch port defined in Access Mode
  - Operating Systems should define OSA without VLAN (VLAN “unaware”)
- Switch port defined in Trunk Mode
  - Operating Systems should define OSA with VLAN (VLAN “aware”)
- An OSA port attaches to a Switch in either Access Mode or Trunk Mode. If multiple LPARs share an OSA port attached to a switch in Access Mode and one of those LPARs wants to start to use VLAN configuration either:
  - Use a second OSA port in Trunk Mode.
  - Or change the Switch port from Access Mode to Trunk Mode. All traffic to/from the LPARs that do not define VLAN will be sent using the Default VLAN ID.
    - If the Default VLAN ID uses a different subnet, then all the LPARs IP addresses will have to change.

OSA MTU, Jumbo Frames, and MTU Discovery

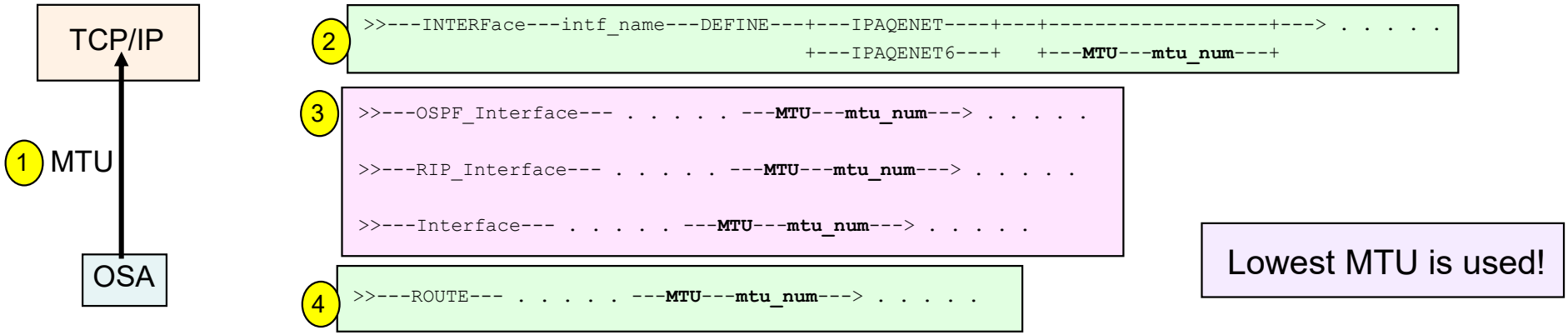



```
>>---INTERFace---intf_name---DEFINE---+---IPAQENET----+---+--------------------+---> . . . . .
                                      +---IPAQENET6---+   +---MTU---mtu_num---+
```

```
                +-------------------------------------------------------------------+
                |                                                                   |
>>---IPCONFig---V---+-----------------------------------------------------------+---+---><
                    |   +---NOPATHMTUDISCovery---+                              |
                    +---+----------------------+-----------------------------------|
                    |   +---PATHMTUDISCovery-----+                              |
                    :                                                           :
                    :                                                           :
```

- MTU mtu_num
  - Specifies the maximum transmission unit (MTU) in bytes.
  - IPCONFig PATHMTUDISCovery may be defined to dynamically discover the path MTU (PMTU), which is the smallest MTU of all the hops in the path. Use this parameter to prevent fragmentation of datagrams.
    - Uses ICMP “fragmentation-needed” errors to detect the PMTU for a path. ICMP errors must be permitted to flow at all hosts along the path of a connection. PATHMTUDISCOVERY does not function if a firewall blocks ICMP errors.
  - When defining Jumbo frames, MTU 8992, **PathMTUDiscovery is recommended**.
  - It is recommended to use the same MTU size on all hosts on the same subnet because there is no router in the path to fragment packets.

OSA Lowest MTU



1. The OSA hardware will notify the TCP/IP stack of the hardware MTU.
2. MTU may be defined on the Interface statement if a lower MTU than the hardware MTU is desired.
   - If MTU is not defined on the Interface, then 2 = 1
3. MTU may be defined on the OMPROUTE definition for the device.
   - If the MTU is not defined on the OMPROUTE definition for the device, the default MTU of 576 is used.
     - Note: Always define all devices in OMPROUTE or use the IGNORE_UNDEFINED_INTERFACE to prevent the default MTU of 576 to be used.
   - The lowest of the three values, 1, 2, or 3, will be used as the MTU by OMPROUTE.
4. MTU may be defined on the ROUTE definition in the BeginRoutes block.
   - If the MTU is not defined on the ROUTE definition, the default MTU of 576 is used.
   - The lowest of the three values, 1, 2, or 4 will be used as the MTU by the routing table.

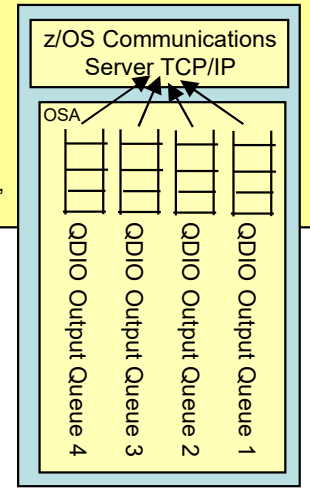
OSA Inbound Blocking and Inbound Workload Queuing

• LCS (non-QDIO OSA OSE mode) and MPCIPA (QDIO OSA OSD mode) LINK

```
>>---LINK---link_name---+---ETHERNet-------+---link_num---device_name--->
                        +---802.3----------+
                        +---ETHEROR802.3---+
                        +---IPAQENET-------+

    +---INBPERF---BALANCED-------------+
>---+----------------------------------+---> . . . . .
    +---INBPERF---+---DYNAMIC------+---+
                  +---MINCPU-------+
                  +---MINLATENCY---+
```

Inbound Workload Queuing (IWQ) INBPERF DYNAMIC WORKLOADQ

IWQ automatically provides unique input queues for:
-Sysplex Distributor traffic
-Bulk data (streaming) traffic
-Enterprise Extender (EE) traffic
-Default (Interactive)
Requires z196+ and z/OS V1.13+
Prevents inbound and outbound out of order packets, and the overhead that goes with it.

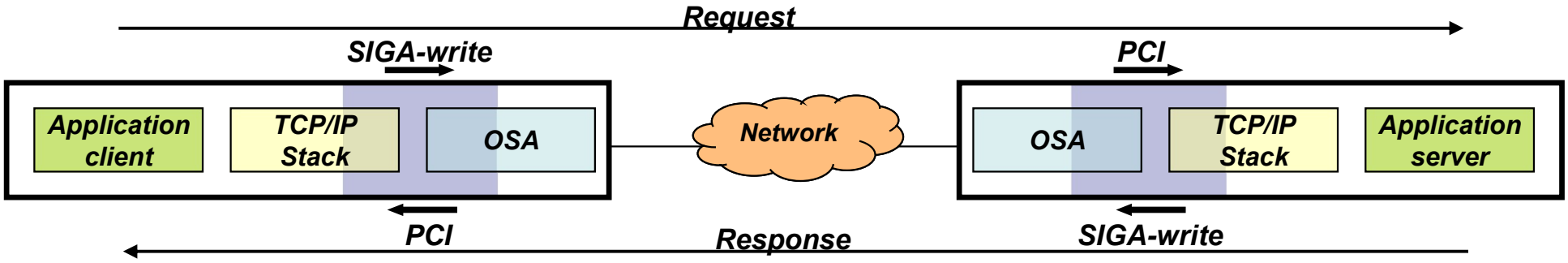



```
                                                    +---INBPERF---BALANCED-----------------+
                                                    +---INBPERF---DYNAMIC---WORKLOADQ------+
>>---INTERFace---intf_name---DEFINE---+---IPAQENET----+---+--------------------------------------+---> . . . . .
                                     +---IPAQENET6---+    +---INBPERF---DYNAMIC---NOWORKLOADQ---+
                                                    +---INBPERF---MINCPU-------------------+
                                                    +---INBPERF---MINLATENCY---------------+
```

- INBPERF
  - Indicates how frequently the adapter should interrupt the host for inbound traffic.
  - 3 Static Settings
    - MINCPU minimizes host interrupts without regard to throughput.
    - MINLATENCY minimizes delay, by more quickly passing packets to the host.
    - BALANCED achieves high throughput and low CPU consumption.
  - 1 Dynamic Setting (z/OS V1.9+, PTFed back to V1.8)
    - DYNAMIC reacts to changes in inbound traffic patterns and sets interrupt-timing values to where throughput is maximized.
    - **DYNAMIC should outperform the other settings for most workload combinations.**
    - See 2098DEVICE Preventive Service Planning (PSP) buckets for hardware support.
    - DYNAMIC WORKLOADQ provides different queues for inbound traffic.
  - INBPERF must match between LINK and INTERFACE for the same OSA.

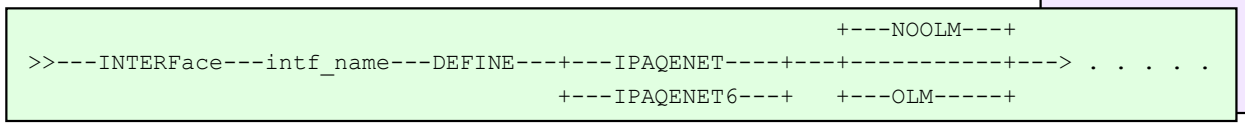
Dynamic LAN Idle Support
LINK . . . INBPERF DYNAMIC
or INTERFACE . . . INBPERF DYNAMIC NOWORKLOADQ

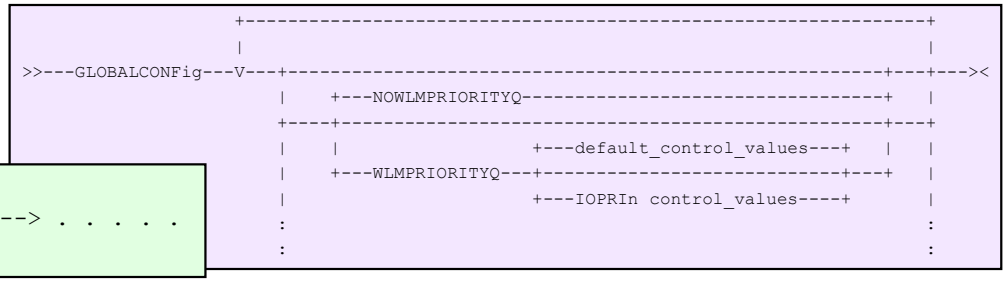
OSA Optimized Latency Mode (OLM)



- Define OLM when,
  - Latency is the most critical factor (ie. More important than CPU overhead).
  - Traffic is not streaming bulk data (ie. FTP).
- Inbound
  - OSA signals the host if data is “on its way” (“Early Interrupt”).
  - Host looks more frequently for data from OSA.
- Outbound
  - OSA does not wait for SIGA to look for outbound data (“SIGA reduction”).

OSA OLM Configuration

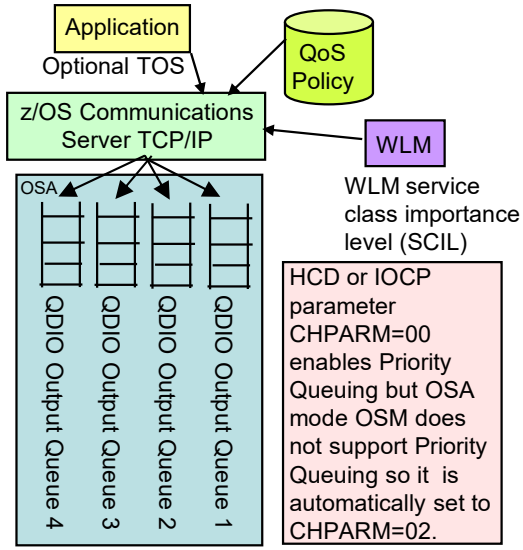




- OLM is specified on the Interface statement.
  - GLOBALCONFIG WLMPRIORITYQ and QoS configuration statement SETSUBNETPRIOTOSMASK may be necessary to benefit from OLM.
    - OLM will not change traffic patterns if all the traffic is being sent to the fourth queue.
- **Restrictions:**
  - **OLM is rarely desired because it is only recommended when Latency is the most critical factor (ie. More important than CPU overhead, etc.)**
  - **When OLM is defined QDIO Accelerator (or HiperSockets Accelerator) will not accelerate the traffic.**
  - **When OLM is specified INBPERF is automatically set to DYNAMIC.**
  - **Interfaces sharing an OSA port using OLM is limited to four.**
    - Each Interface statement counts toward the 4 Interface limit:
      - LPAR TCP/IP stack using the OSA port
      - VLAN defined for this OSA port
      - Protocol (IPv4 or IPv6) interface defined for this OSA port
      - TCP/IP stack on the same LPAR using the OSA port
      - TCP/IP stack activating the OSA-E Network Traffic Analyzer (OSAENTA)

OSA Outbound Priority Routing Queues

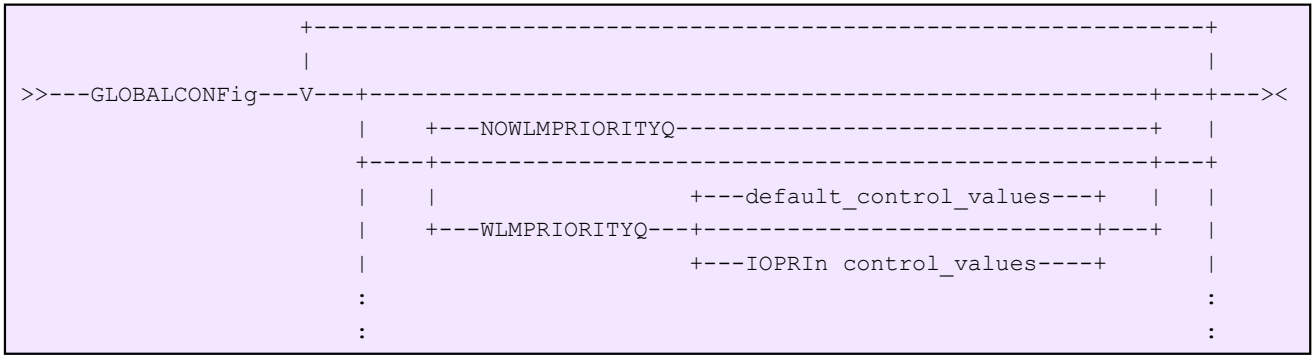
- The Type of Service (ToS) byte in the IP header may be used by routers in the IP network to prioritize traffic (forward some types of traffic before others).
  - The most benefit is realized when the routers are all configured for this support.
- TCP/IP uses the first three bits of the ToS byte in the IP header to determine the outbound priority value for a given datagram.
  - Optionally an application can specify the TOS for its traffic.
- z/OS CS TCP/IP supports four priority values in the range 1–4 for outbound QDIO traffic (with 1 being the highest priority).
  - TCP/IP will send packets using these four queues whether or not any routers in the network are configured to use the ToS settings.
- z/OS CS TCP/IP Policy Agent Quality of Service (QoS) may be used to override the default mapping of ToS values to priorities.
  - This may be used for devices without VLANs.
    - SetSubnetPrioTosMask statement
  - This may be used for devices with VLANs.
    - PriorityTosMapping parameter on the SetSubnetPrioTosMask statement may define VLAN priority-tagging.
- Enterprise Extender (EE) (SNA encapsulation over IP) automatically configures IP ToS.



Default mapping of ToS values to priorities:

ToS	Priority
000	4
001	4
010	3
011	2
100	1
101	1
110	1
111	1

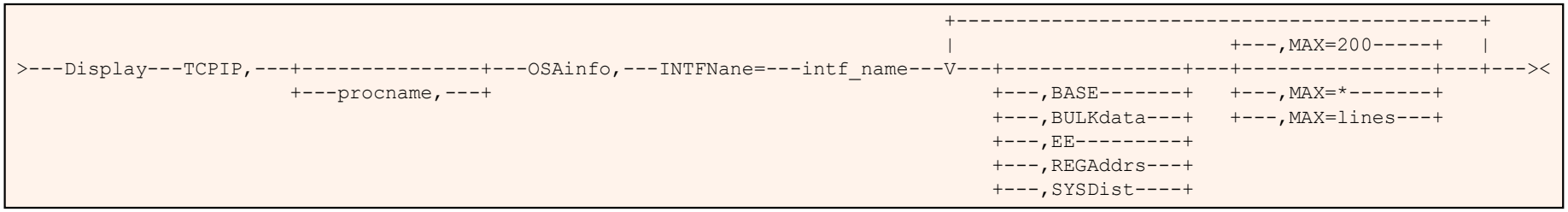
WLM Service Class for OSA Queuing



- WLM IO Priority Enhancement
  - When the GLOBALCONFIG WLMPRIORITYQ parameter is specified and a packet with a ToS or traffic class value 0 is sent over QDIO OSA port, TCP/IP sets the QDIO write priority of the packet based on the priority value provided by the WLM service class.

WLM Service classes	TCP/IP assigned control value	Default QDIO queue mapping
SYSTEM	n/a	Always queue 1
SYSSTC	0	Queue 1
User-defined with IL 1	1	Queue 2
User-defined with IL 2	2	Queue 3
User-defined with IL 3	3	Queue 3
User-defined with IL 4	4	Queue 4
User-defined with IL 5	5	Queue 4
User-defined with discretionary goal	6	Queue 4

Query and Display OSA Configuration



- OSAINFO displays the current OSA configuration.
  - Displays all sections when no filters are specified.
  - BASE displays physical characteristics and attributes for the interface.
  - BULKDATA displays IWQ routing variables for the BULKDATA ancillary queue.
    - BULKDATA routing variables are source and destination IP addresses, source and destination ports, and protocol.
  - EE displays IWQ routing variables for the Enterprise Extender (EE) ancillary queue.
    - EE routing variables are destination IP addresses, destination ports, and protocol.
  - REGADDRS displays registered Layer 3 unicast and multicast addresses.
  - SYSDIST displays IWQ routing variables for Sysplex Distributor ancillary queue.
    - SYSDIST routing variables are destination IP addresses and protocol.
- Background
  - Prior to OSAINFO, OSA/SF was often used for OSA QDIO to retrieve active information.

VTAM QDIOSYNC Overview

- VTAM QDIOSYNC trace captures OSA diagnostic data.
- Instead of or in addition to using the HMC (Hardware Management Console) to manually capture OSA diagnostic data, QDIOSYNC may be used to cause the OSA to automatically capture diagnostic data when:
  - OSA detects an unexpected loss of host connectivity.
    - Unexpected halt signal from host
    - Host unresponsive
  - OSA receives CAPTURE signal from host due to:
    - VTAM-supplied MPF (Message Processing Facility) exit (IUTLLCMP) is driven.
      - Add VTAM-supplied MPF exit module, USEREXIT(IUTLLCMP), to SYS1.PARMLIB(MPFLISTxx).
      - Issue SET MPF=(xx,zz) where xx is the new PARMLIB member and zz is the old.
        - » Activates the new MPFLISTxx member.
      - Set corresponding SLIP trap to initiate a host dump.
      - See z/OS MVS Installation Exits for more information about MPF.
    - VTAM or TCP/IP FRR (Functional Recovery Routine) is driven with ABEND06F.
      - Result of SLIP PER trap that specifies ACTION=RECOVERY.
- After QDIOSYNC trace use the HMC to copy the OSA diagnostic data.

Start Option or Modify Command

- VTAM start options and commands
  - MODIFY TRACE and NOTRACE with TYPE=QDIOSYNC
    - Activate and terminate QDIOSYNC trace.
    - ID=trle_name
      - To activate that trace on a single OSA.
    - ID=* is supported
      - SAVE=NO applies QDIOSYNC trace to all currently active OSA TRLEs.
      - SAVE=YES applies QDIOSYNC trace to all currently active and future active OSA TRLEs.
- QDIOSYNC TRACE OPTION filters
  - ALLIN
    - Collects only inbound diagnostic data for all OSAs.
  - ALLOUT
    - Collects only outbound diagnostic data for all OSAs.
  - ALLINOUT
    - Collects inbound and outbound diagnostic data for all OSAs.
  - IN
    - Collects only inbound diagnostic data only for OSAs defined to this VTAM.
  - OUT
    - Collects only outbound diagnostic data only for OSAs defined to this VTAM.
  - INOUT
    - Collects inbound and outbound diagnostic data only for OSAs defined to this VTAM.

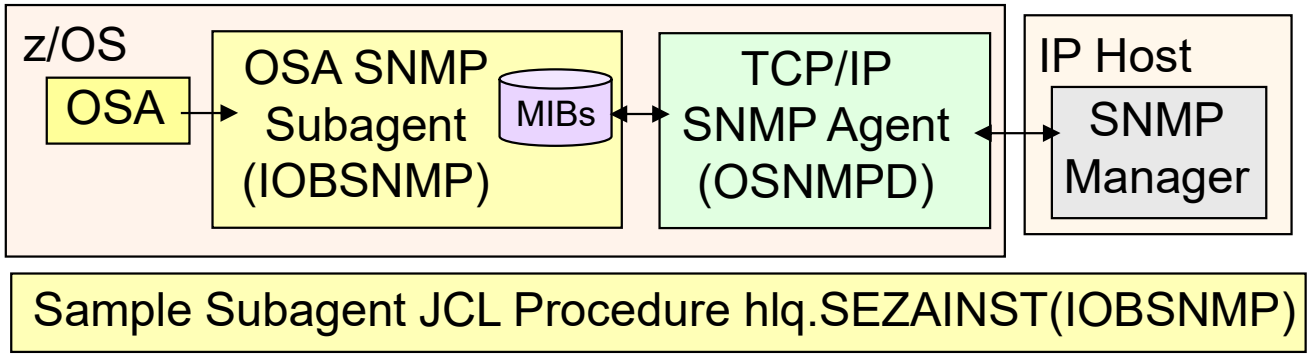
OSAENTA Overview

- OSA-Express network traffic analyzer (OSAENTA) traces frames for an OSA in QDIO mode.
  - Also known as OSA Network Traffic Analyzer (OSA NTA)
- OSAENTA is controlled and formatted by z/OS Communications Server (CS), but is collected in the OSA port.
  - Support Element (SE) customization is required to enable or disable OSAENTA and authorize tracing outside the local LPAR.
- OSAENTA has capabilities beyond other z/OS CS tracing:
  - Trace frames discarded by the OSA.
  - Trace the MAC headers for packets.
  - Trace ARP packets
  - Trace packets to and from other users sharing the OSA (other TCP/IP stacks, z/Linux users, and z/VM users)
  - Trace SNA packets

Update Profile or Vary TCPIP

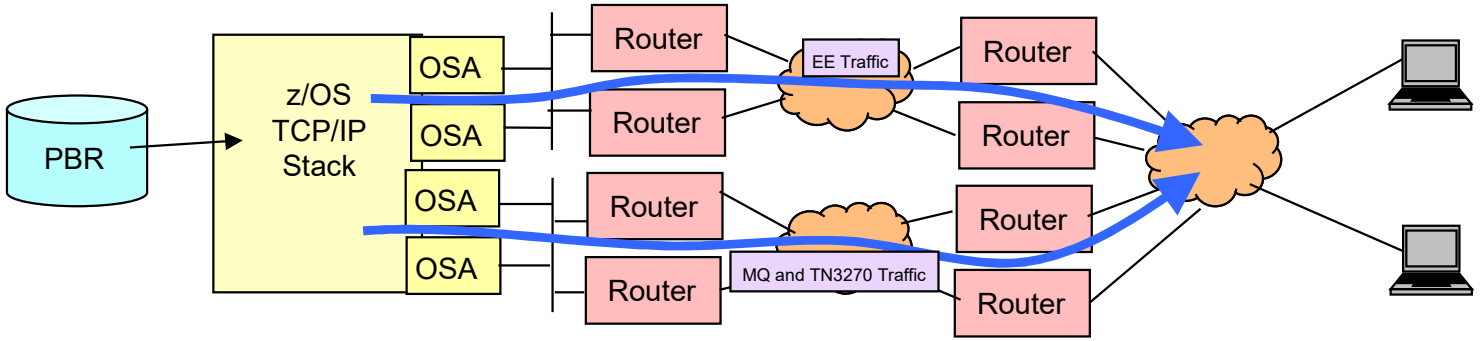
- Control OSAENTA trace using OSAENTA statement in the TCP/IP Profile or the VARY TCPIP,,OSAENTA command.
  - Filter what data is collected
    - IP address
    - Protocol (TCP,UDP,etc.)
    - Port number
    - Frame type
    - There is a limit of only one filter value per OSAENTA statement/command.
    - There is a limit of up to 8 filter values per filter (ie. only 8 port numbers may be defined).
    - Up to 8 IPv4 addresses and up to 8 IPv6 addresses may be specified.
    - All frames that match any IP address and match all other filters are captured.
  - Specify how much data is to be collected.
  - OSAENTA command OPERCMDS resource name is MVS.VARY.TCPIP.OSAENTA.
- Display current OSAENTA trace settings using the Netstat DEvlinks/-d command.
- OSAENTA dynamically creates interface EZANTAxxxxxxxx.
  - xxxxxxxx is the port name in the OSAENTA command and the TRLE.
    - TRLE must exist.
  - Used for receiving trace records.
  - Use VARY TCPIP,,OSAENTA commands ON, OFF, and DEL to start, stop, and delete the OSAENTA interface.
- CTRACE (Component Trace) uses SYSTCPOT to collect the trace records.
  - IPCS CTRACE with component name SYSTCPOT may be used to format the trace.

OSA SNMP Support



- OSA provides an SNMP Subagent (IOBSNMP)
  - Available for use with TCP/IP SNMP Agent (OSNMPD)

Outbound Routing



- Policy-based Routing (PBR) of Outbound Traffic (traffic that Originates on z/OS)
  - Choose first hop router, outbound network interface (including VLAN), and MTU
  - Choice can be based on more than the usual destination IP address/subnet
  - With PBR, the choice can be based on source/destination IP addresses, source/destination ports, TCP/UDP, etc.
  - Allows an installation to separate outbound traffic for specific applications to specific network interfaces and first-hop routers:
    - Security related
    - Choice of network provider
    - Isolation of certain applications
      - EE traffic over one interface
      - TN3270 traffic over another interface
  - PBR policies will identify one or more routes to use
    - If none of the routes are available, options to use any available route or to discard the traffic will be provided

Some Useful Commands for OSA Information

- IP Commands
  - See the IP System Administrator's Commands manual for syntax and details.
  - Vary TCPIP,procname,OSAENTA
    - Control the OSA-Express Network Traffic Analyzer (OSAENTA) tracing facility.
  - Vary TCPIP,procname,START/STOP,device_name/interface_name
    - Start or stop device or interface.
  - NETSTAT ARp/-R ip_addr/ALL…
    - Output includes the ip_addr to MAC address mapping.
  - NETSTAT DEvlinks/-d…
    - Output includes information about devices, links, and interfaces.
  - NETSTAT HOme/-h…
    - Output includes the IP Addresses to links/interfaces mapping.
  - NETSTAT SRCIP/-J…
    - Output includes Source IP Address information.
- VTAM Commands
  - See the SNA Operations manual for syntax and details.
  - DISPLAY NET,ID=xcaname…
    - Output includes LINE and PU.
  - DISPLAY NET,TNSTAT
    - Output indicates which TRLEs are collecting statistics if start option TNSTAT is specified.
  - DISPLAY NET,TRL… and DISPLAY NET,TRL,TRLE=…
    - Output includes TRL/TRLE information.
  - MODIFY procname,TRACE/NOTRACE,TYPE=QDIOSYNC…
    - Control QDIO Synchronization Trace