z/OS V2R2 Communications Server Performance Summary

Dan Patel
danpatel@us.ibm.com
Dave Herr
dherr@us.ibm.com

Contents

Contents (cont'd)

Trademarks, notices, and disclaimers

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both:

- Advanced Peer-to-Peer Networking®
- AIX®
- alphaWorks®
- AnyNet®
- AS/400®
- BladeCenter®
- Candle®
- CICS®
- DataPower®
- DB2 Connect
- DB2®
- DRDA®
- e-business on demand®
- e-business (logo)
- e business(logo)®
- ESCON®
- FICON®
- GDDM®
- GDPS®
- Geographically Dispersed Parallel Sysplex
- HiperSockets
- HPR Channel Connectivity
- HyperSwap
- i5/OS (logo)
- i5/OS®
- IBM eServer
- IBM (logo)®
- IBM®
- IBM zEnterprise™ System
- IMS
- InfiniBand ®
- IP PrintWay
- IPDS
- iSeries
- LANDP®
- Language Environment®
- MQSeries®
- MVS
- NetView®
- OMEGAMON®
- Open Power
- OpenPower
- Operating System/2®
- Operating System/400®
- OS/2®
- OS/390®
- OS/400®
- Parallel Sysplex®
- POWER®
- POWER7®
- PowerVM
- PR/SM
- pSeries®
- RACF®
- Rational Suite®
- Rational®
- Redbooks
- Redbooks (logo)
- Sysplex Timer®
- System i5
- System p5
- System x®
- System z®
- System z9®
- System z10
- Tivoli (logo)®
- Tivoli®
- VTAM®
- WebSphere®
- xSeries®
- z9®
- z10 BC
- z10 EC
- zEnterprise
- zSeries®
- z/Architecture
- z/OS®
- z/VM®
- z/VSE

* All other products may be trademarks or registered trademarks of their respective companies.

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both:

- Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
- Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.
- Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
- Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
- InfiniBand is a trademark and service mark of the InfiniBand Trade Association.
- Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
- ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
- IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

Notes:

- Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
- IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
- All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
- This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
- All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
- Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the 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.
- Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Refer to www.ibm.com/legal/us for further legal information.

V2R2 performance and scalability functions

- Shared Memory Communications
  - Shared Memory Communications – Direct Memory Access (SMC-D)
    - New in V2R2
      Improved SMC-R sending queued data algorithm
  - Shared Memory Communications – RDMA (SMC-R)
    - Improved SMC-R sending queued data algorithm
      - Send up to three queued writes in one interrupt
      - Benefit streaming/bulk type workloads
    - Default to use 1MB (largest) RMBe for streaming/bulk data receive side
    - FTP uses 180K receive buffer
    - SMC-R virtualization – LPARs sharing a RoCE Express feature

V2R2 performance and scalability functions …

- New/improved TCP/IP autonomics
  - Avoid DELAYACK timer processing
    - AUTODELAYACK option
    - Eliminate occasional delays
  - Dynamic Right Sizing (DRS) improvement
    - Remain enabled
    - Unlimited runway to enablement
  - Outbound “Dynamic Right Sizing (ORS)”
    - Allow send buffer to grow when sending streaming (bulk) data
  - Improved response to lost vs. out-of-order packets
    - Don’t reduce slow start threshold for out-of-order
  - VIPAROUTE fragmentation avoidance
    - GLOBALCONFIG parameter – AUTOADJUSTMSS

V2R2 performance and scalability functions …

- 64-bit Enablement of TCPIP stack and DLCs
- VTAM Internal Trace Improvement and Trace Recommendation
- Enhance IKE Scalability
- Enhanced Enterprise Extender (EE) scalability
  - Large number EE connections (thousands)
  - Improved caching – improved latency and cpu

Shared Memory Communications SMC-D architecture

Faster communications that preserve TCP/IP qualities of service

- **Shared Memory Communications – Direct Memory Access (SMC-D) optimizes z/OS for improved performance in '*within-the-box*' communications versus standard TCP/IP over HiperSockets or Open System Adapter**

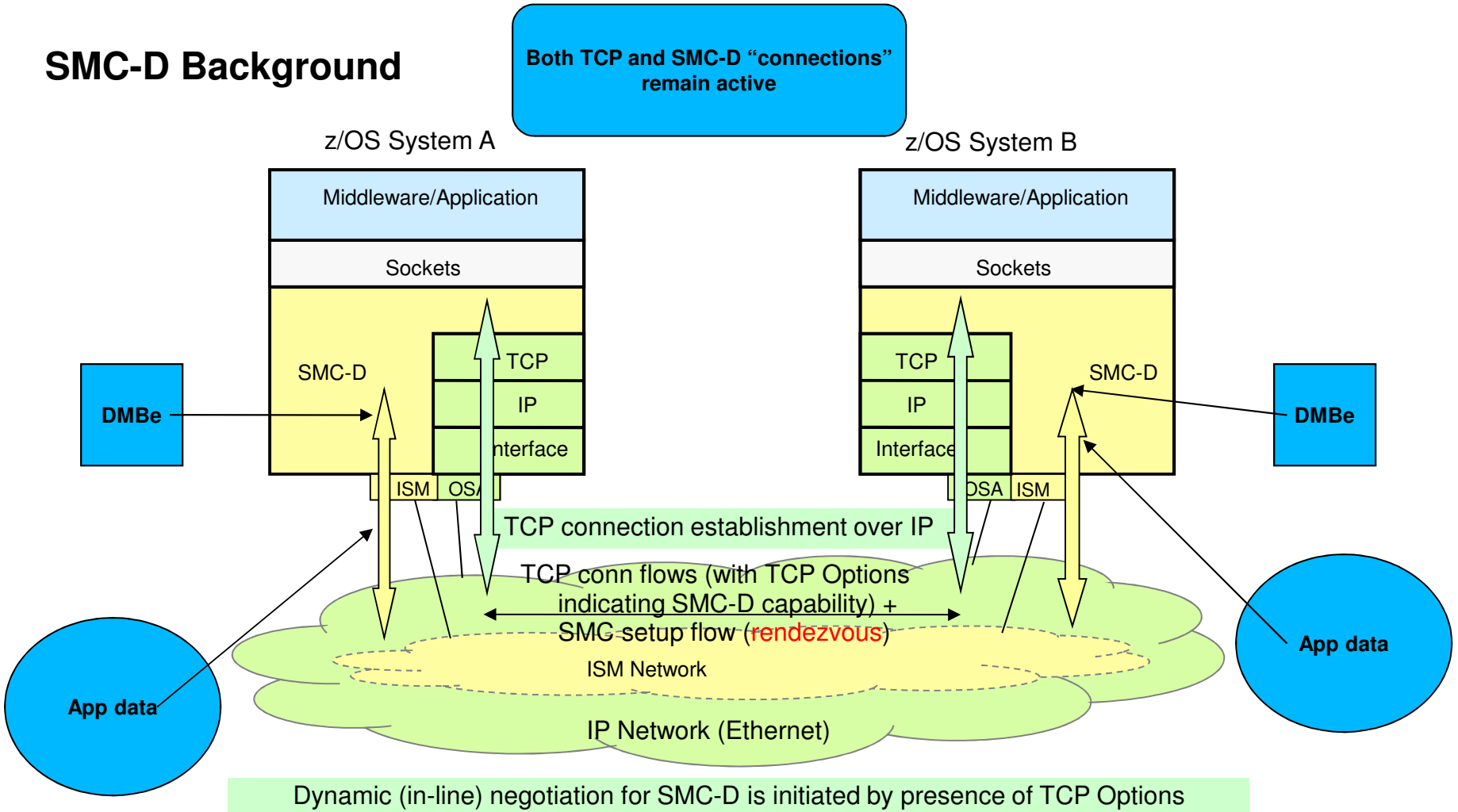
Typical Client Use Cases:

- **Valuable for multi-tiered work co-located onto a single z Systems server without requiring extra hardware**
- **Any z/OS TCP sockets based workload can seamlessly use SMC-D without requiring any application changes**

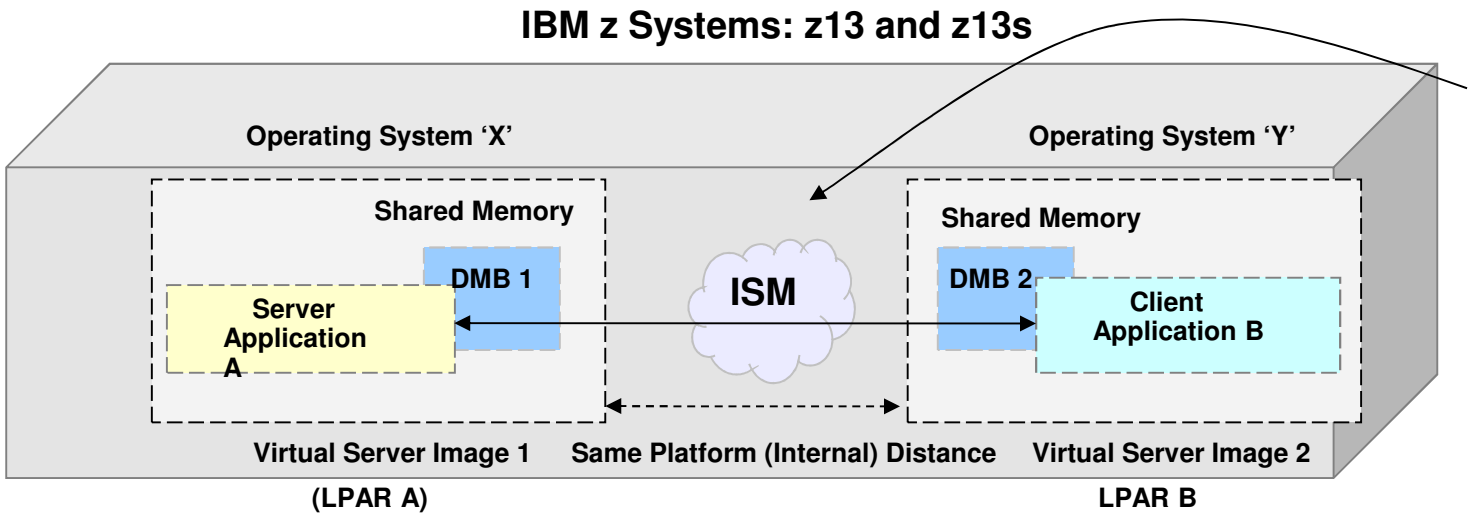
SMC Applicability Tool (SMCAT) is available to assist in gaining additional insight into the applicability of SMC-D (and SMC-R) for your environment

Shared Memory Communications – SMC-D

SMC-D Background



Shared Memory Communications-Direct Memory Access (SMC-D) over Internal Shared Memory (ISM)



"Shared Memory"

across unique OS instances within the same CPC

- SMC-D (over ISM) extends the value of the Shared Memory Communications architecture by enabling SMC for direct LPAR to LPAR communications. SMC-D is very similar to SMC-R extending the benefits of SMC-R to same CPC operating system instances without requiring physical resources (RoCE adapters, PCI bandwidth, NIC ports, I/O slots, network resources, 10GbE switches etc.).
- Eliminates TCP/IP processing in the data path.

Note 1. The performance benefits of SMC-R (cross CPC) and HiperSockets (within CPC) are similar to each other. SMC-D / ISM provides significantly improved performance benefits above both within the CPC.

ISM System z13 and z/OS SMC-D Requirements

1. IBM z Systems: z13 (driver level 27 (GA2)) or z13s
2. z/OS software (PTF) requirements:

    1. CommServer VTAM: OA48411 UA80711
    2. CommServer TCP/IP: PI45028 UI35411
    3. z/OS (IOS): OA47913 UA80812
    4. HCD: OA46010
    5. IOCP: OA47938 UA90986
    6. HCM: IO23612
    7. RMF OA49113 UA80445

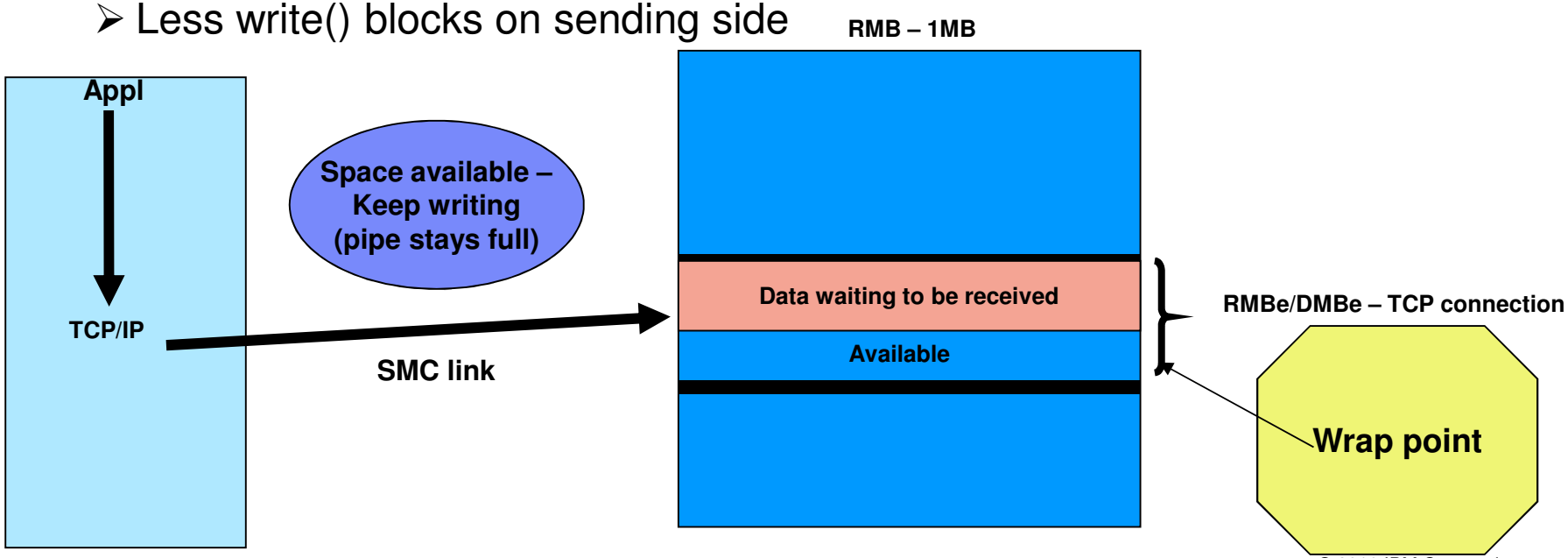
    Note: For a complete / current list of PTFs refer to the PSP bucket.

- For latest info on SMC-R, SMC-D and SMCAT :

http://www-01.ibm.com/software/network/commserver/SMC/

SMC-R/SMC-D – Role of the RMBe/DMBe (buffer size)

- The RMBe/DMBe is a slot in the RMB/DMB buffer for a specific TCP connection
  - Based on, not necessarily equal to, TCPRCVBufrsize
  - Can be controlled by application using setsockopt() SO_RCVBUF
  - 5 sizes - 32K (SMC-R only), 64K, 128K, 256K and 1024K (1MB)
  - Will use 1MB for TCPRCVBufrsize > 128K
  - Depending on the workload, a larger RMBe/DMBe can improve performance
    - Streaming (bulk) workloads
    - Less wrapping of the RMBe/DMBe = less RDMA writes
    - Less frequent “acknowledgement” interrupts to sending side
    - Less write() blocks on sending side



System z13 SMC-D Overall Performance Summary

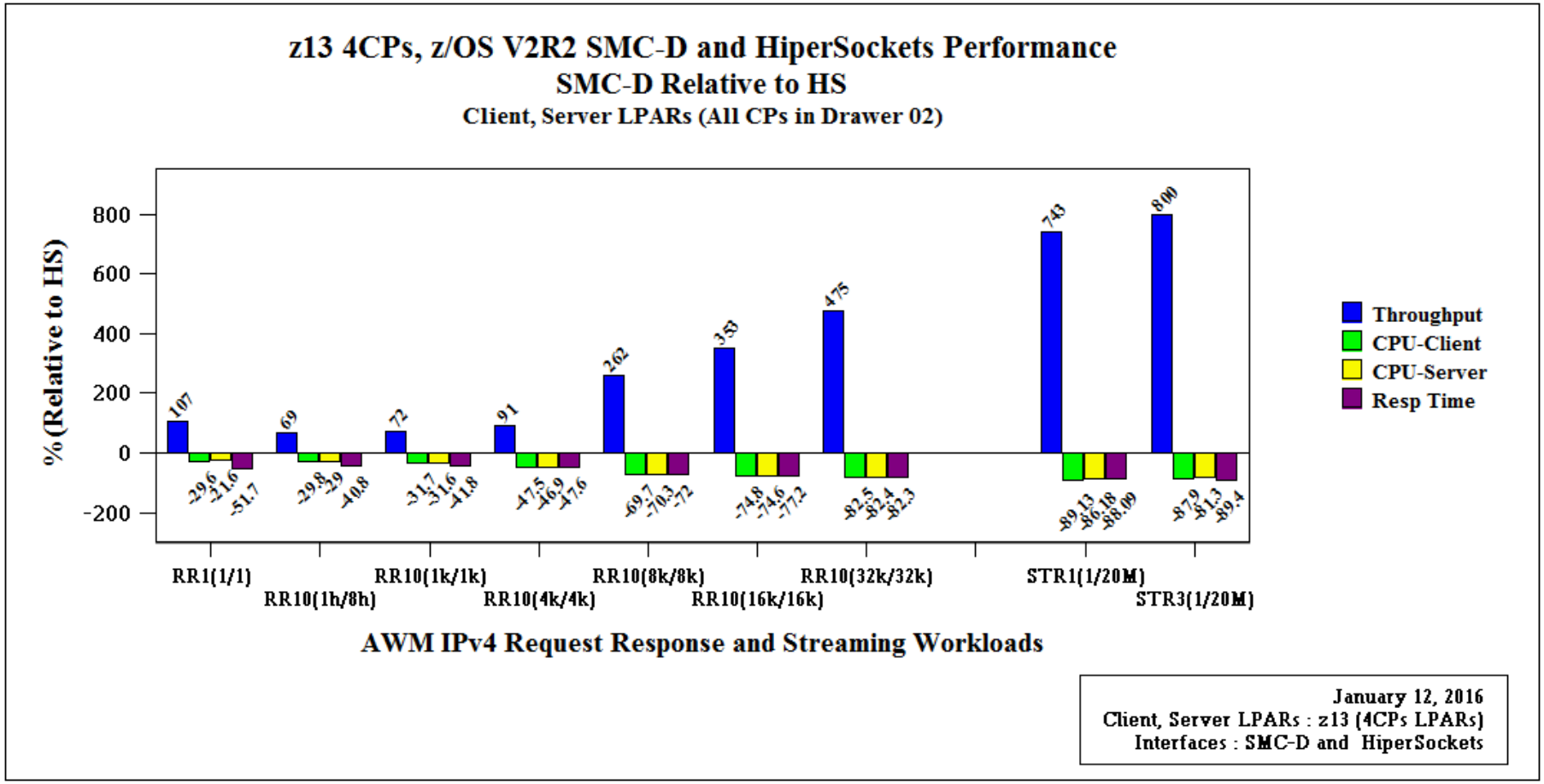
- Performance results are based on IBM Internal micro benchmarks using standard tools used for z/OS release.
  - Environment :
    - Setup used : z3 4CPs Client, Server LPARs using same drawer with V2R2 Communications Server includes latest software and GA2 system firmware.
  - All Results Compared to HiperSockets using 8k frame size.
    - Request/Response Workloads: For 4k/4k workload Up to 48% reduction in latency , Up to 91% increase in throughput, while realizing up to 47% reduction in network related CPU cost. Note, for higher Payloads 8k/8k…32k/32k (72 - 82)% reduction in latency, (262 - 475)% increase in throughput and (70 to 82)% reduction in network related CPU cost.
    - Streaming Workload: For Streaming workloads, (79-89)% reduction in latency, (369 – 789)% increase in throughput, while realizing (80 – 87)% reduction in network related CPU cost
    - FTP for Binary Get and Put observed up to 60% lower CPU cost and equivalent throughput. Note: For FTP transfers the throughput is gated by DASD I/O . Here we used High Performance FICON DASD and 1200 MB MVS dataset and used write / read to DASD.

SMC-D / ISM Performance Details

- The following Charts provide additional SMC-D performance details with comparison of SMC-D / ISM to:
    1. HiperSockets
    2. OSA
    3. SMC-R

    .

HiperSockets Comparison



Up to 9x the throughput! See breakout summary on next chart.

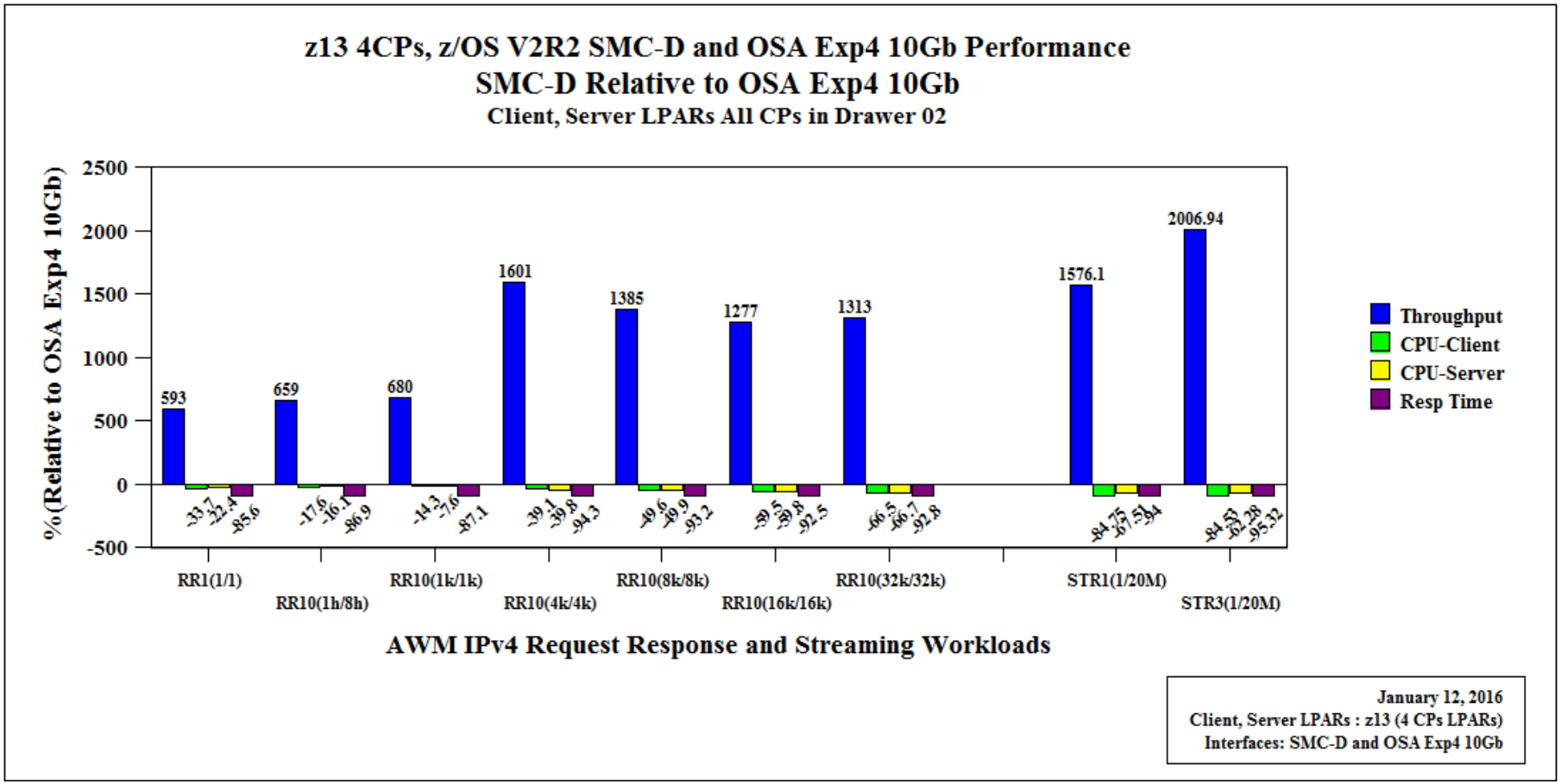
SMC-D / ISM to HiperSockets Summary Highlights

- **Request/Response Summary for Workloads with 1k/1k – 4k/4k Payloads:**
  - **Latency: Up to 48% reduction in latency**
  - **Throughput: Up to 91% increase in throughput**
  - **CPU cost: Up to 47% reduction in network related CPU cost**

- **Request/Response Summary for Workloads with 8k/8k – 32k/32k Payloads:**
  - **Latency: Up to 82% reduction in latency**
  - **Throughput: Up to 475% (~6x) increase in throughput**
  - **CPU cost: Up to 82% reduction in network related CPU cost**

- **Streaming Workload:**
  - **Latency: Up to 89% reduction in latency**
  - **Throughput: Up to 800% (~9x) increase in throughput**
  - **CPU cost: Up to 89% reduction in network related CPU cost**

OSA Comparison



Up to 21x the throughput! See breakout summary on next chart.

SMC-D / ISM to OSA Summary Highlights

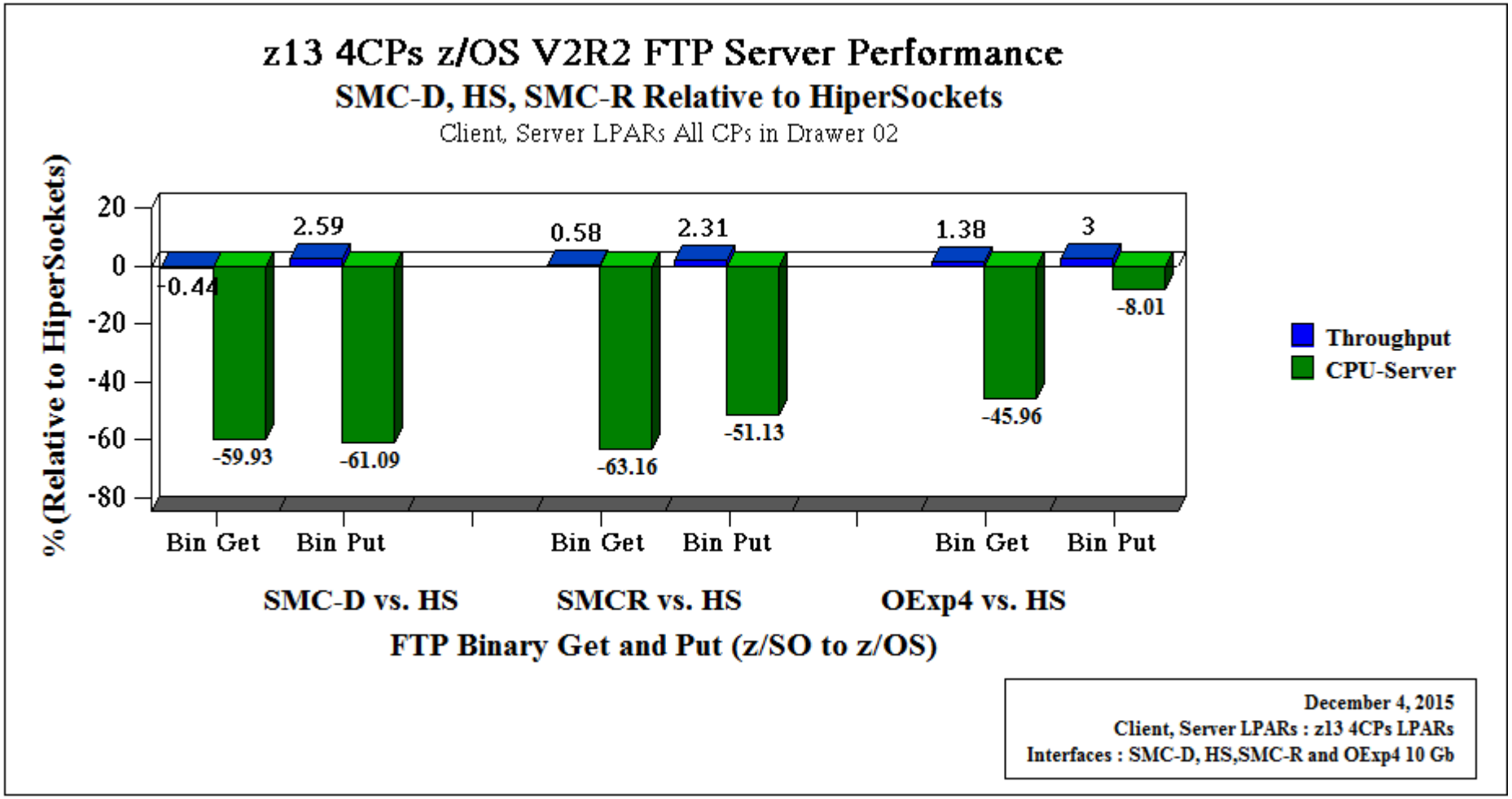
- **Request/Response Summary for Workloads with 1k/1k – 4k/4k Payloads:**
  - **Latency: Up to 94% reduction in latency**
  - **Throughput: Up to 1601% (~17x) increase in throughput**
  - **CPU cost: Up to 40% reduction in network related CPU cost**

- **Request/Response Summary for Workloads with 8k/8k – 32k/32k Payloads:**
  - **Latency: Up to 93% reduction in latency**
  - **Throughput: Up to 1313% (~14x) increase in throughput**
  - **CPU cost: Up to 67% reduction in network related CPU cost**

- **Streaming Workload:**
  - **Latency: Up to 95% reduction in latency**
  - **Throughput: Up to 2001% (~21x) increase in throughput**
  - **CPU cost: Up to 85% reduction in network related CPU cost**
- **FTP:**
  - **For Binary Get and Put:**
    - **Up to 58% lower (receive side) CPU cost and**
    - **Up to 26% lower (send side) CPU cost and equivalent throughput**

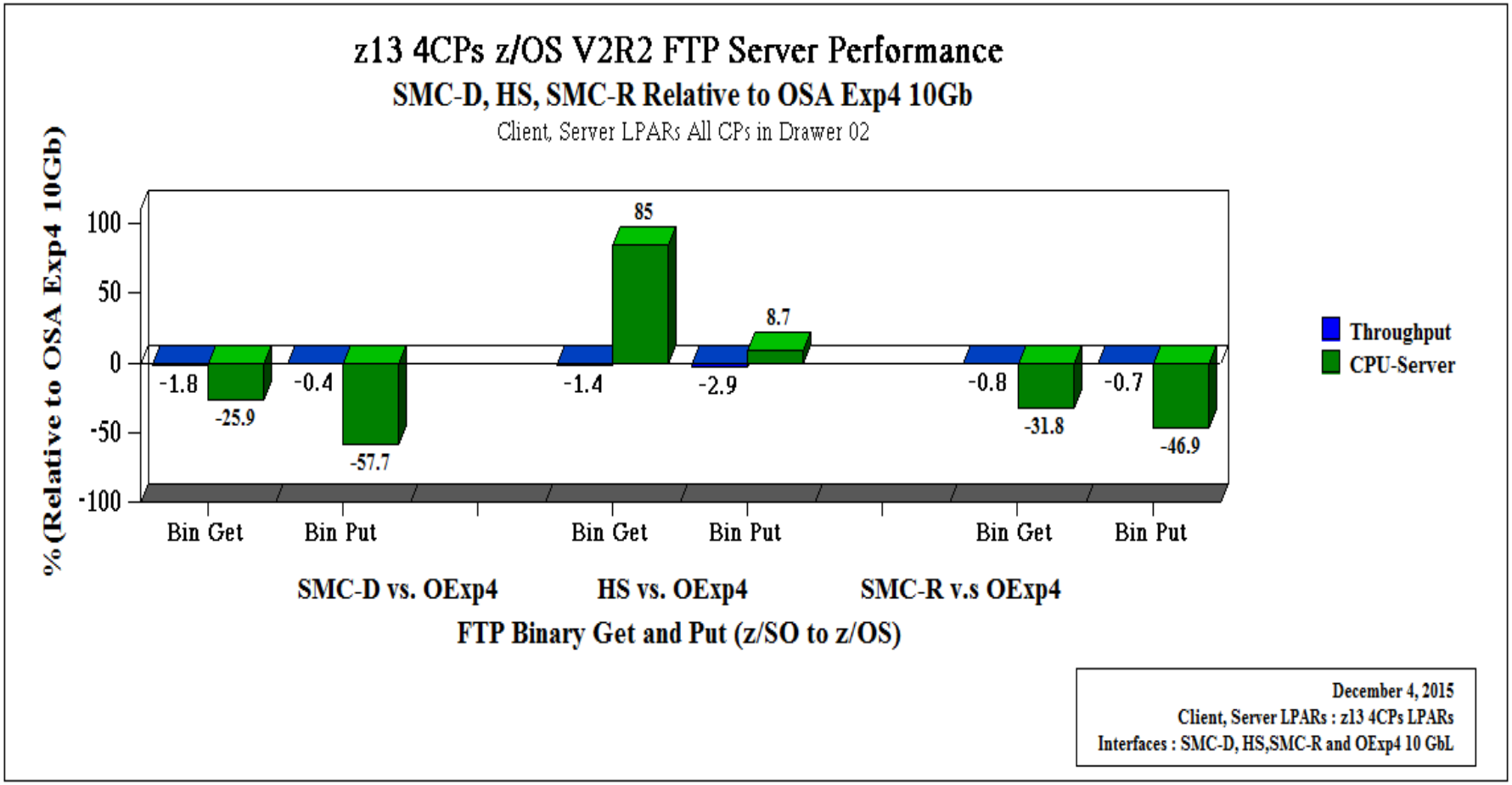
FTP Performance Comparison Relative to HiperSockets

FTP Performance : SMC-D, SMC-R and OSA Relative to HiperSockets



FTP Performance Relative to OSA

- SMC-D provides significant CPU benefits in comparison with OSA, SMC-R and HS
- Throughput is constrained by I/O in all measurements



FTP using zHPF – Improving throughput

- There are many factors that influence the transfer rates for z/OS FTP connections. Some of the more significant ones are (in order of impact):
  - **DASD read/write access**
  - Data transfer type (Binary, ASCII..)
  - Dataset characteristics (e.g., fixed block or variable)

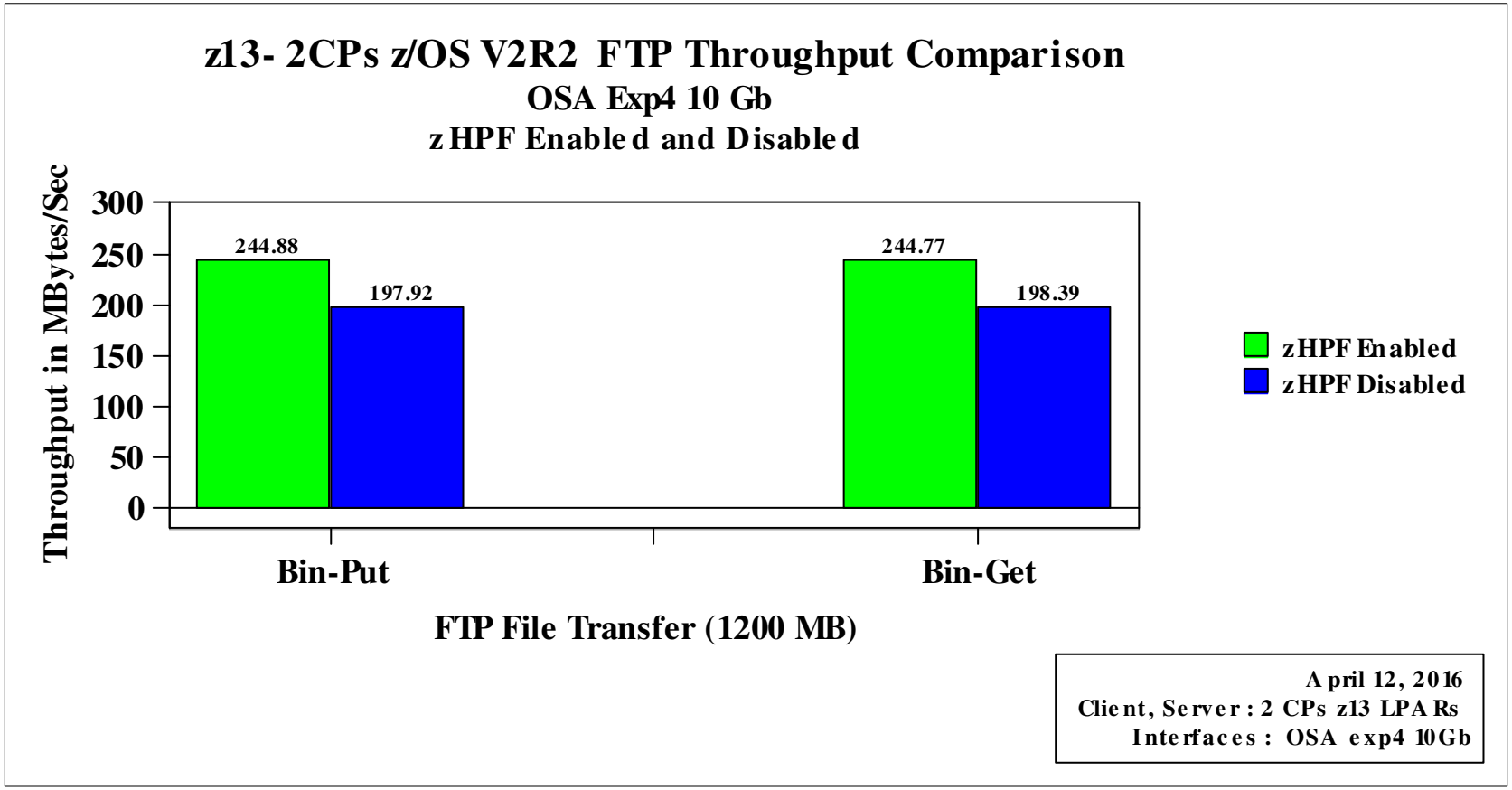
  *Note the network (HiperSockets, OSA, 10Gb, SMC-R) characteristics have very little impact when reading from, and writing to, DASD as you will see in our results section.
- zHPF FAQ link
  - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FQ127122
  - Works with DS8000 storage systems
  - IBM System z High Performance FICON technology improves FTP performance significantly.

FTP using zHPF – Improving throughput

- FTP Workload
  - z/OS FTP client GET or PUT 1200 MB data set from or to z/OS FTP server
  - DASD to DASD (read from or write to)
  - zHPF enabled/disabled
  - Single file transfer
  - Used Variable block data set for the test
    - Organization .... PS
    - Record Format ...VB
    - Record Length …6140
    - Block size 23424
  - For Hipersocket
    - Configure GLOBALCONFIG IQDMULTIWRITE

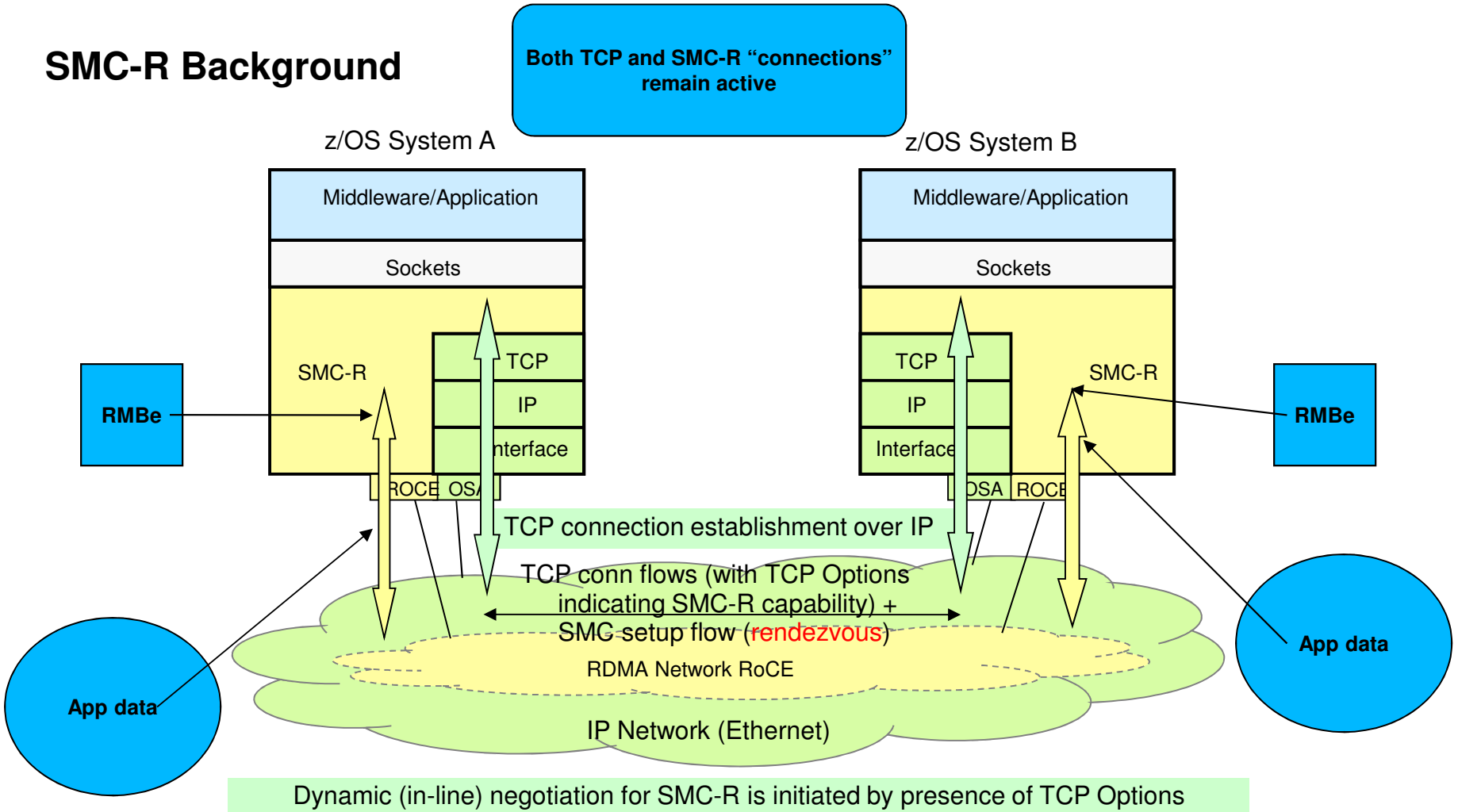
FTP using zHPF – Improving throughput

Throughput is improved by 23-24% with Enabling zHPF



Shared Memory Communications – Remote (SMC-R)

SMC-R Background

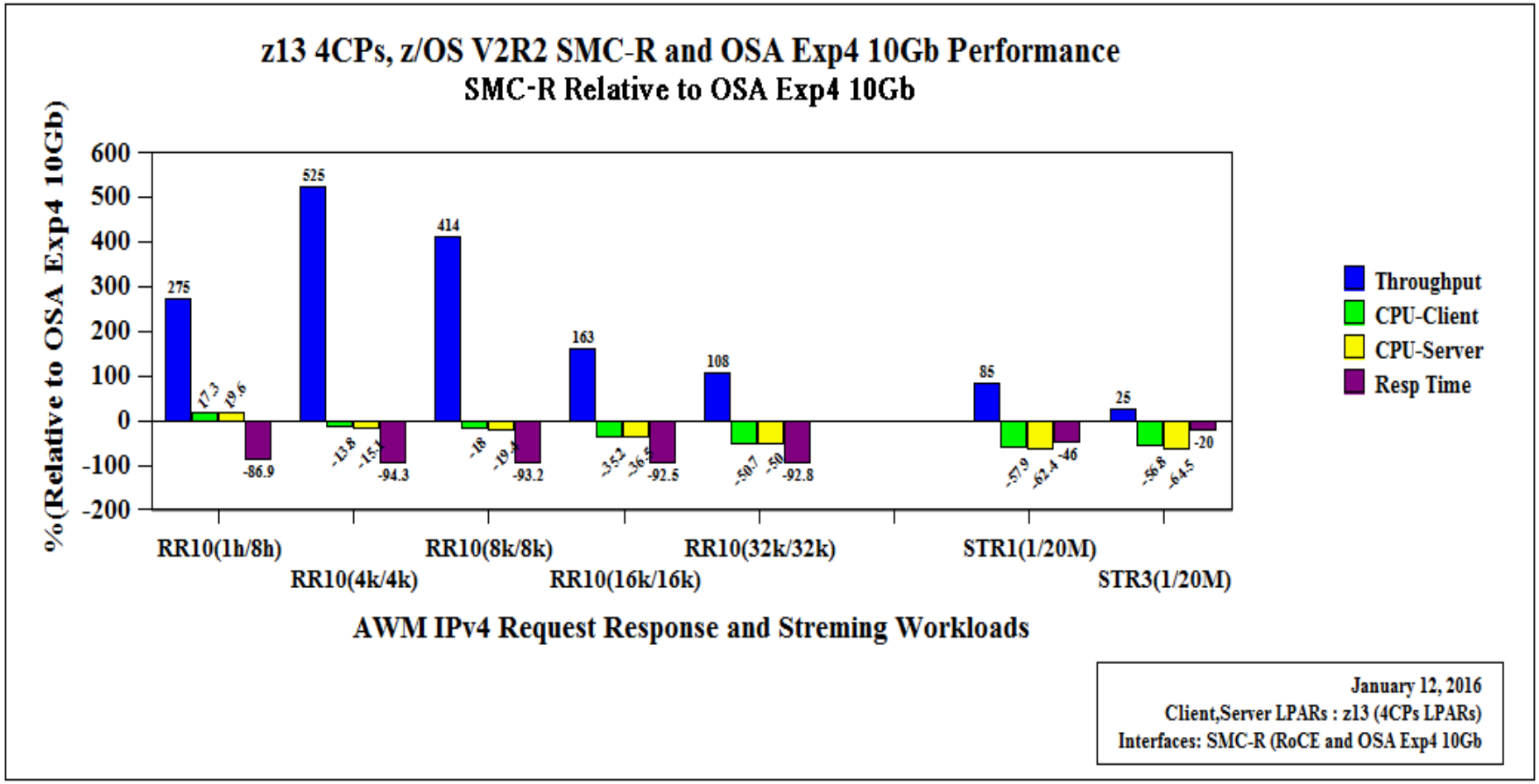


Dynamic (in-line) negotiation for SMC-R is initiated by presence of TCP Options

TCP connection transitions to SMC-R allowing application data to be exchanged using RDMA

SMC-R Performance Relative to OSA

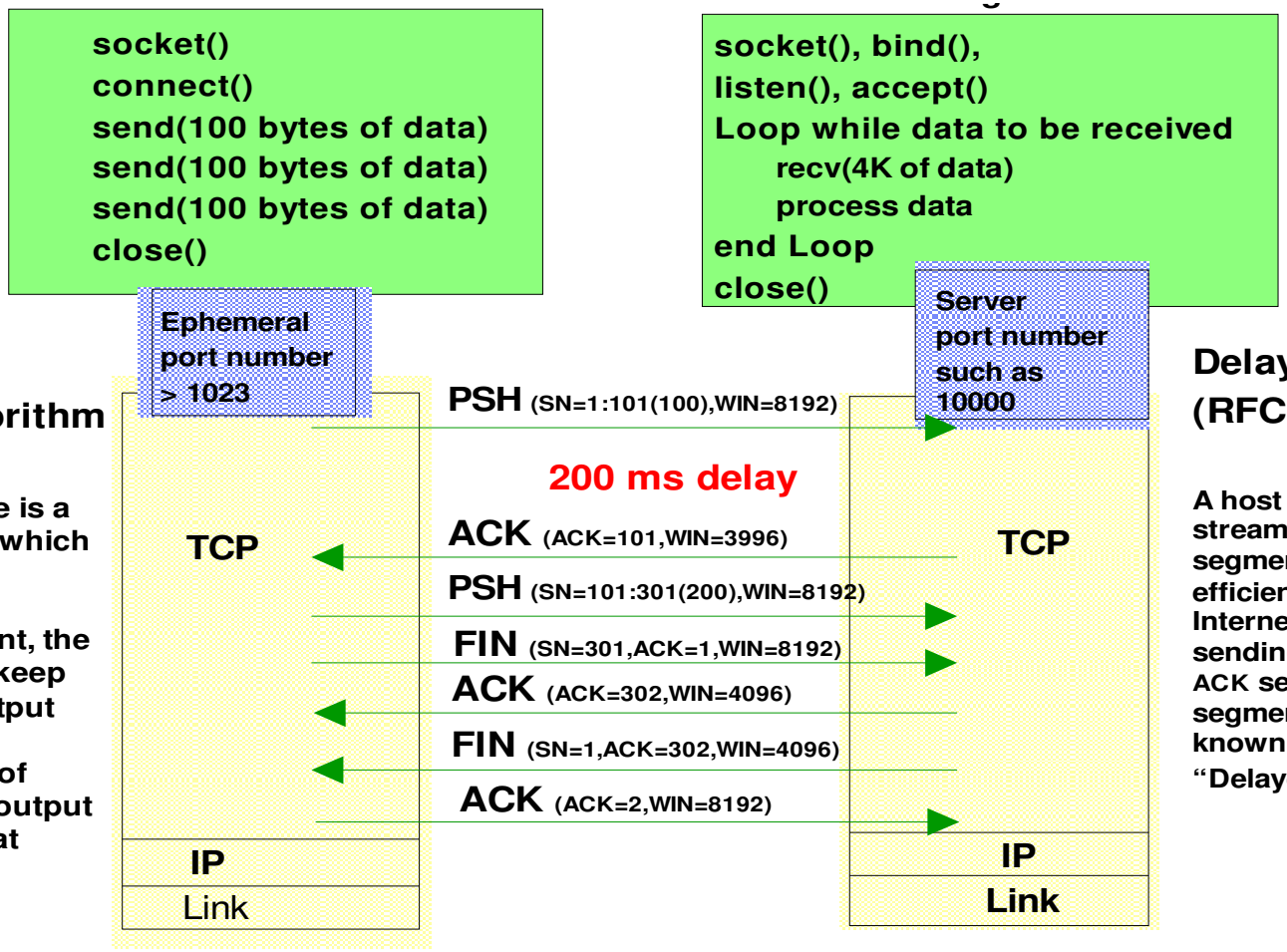
SMC-R Performance Relative to OSA Exp4/5 10Gb



- Up to 6x the throughput for Request Response workload and 50% lower CPU cost
- Up to 64% lower CPU cost for Streaming workload

TCP Delayed Ack Processing

TCP Delayed Ack Processing and Nagle's Algorithm "Catch-22"



Nagle's Algorithm (RFC896)

As long as there is a sent packet for which the sender has received no acknowledgment, the sender should keep buffering its output until it has a full packet's worth of output, so that output can be sent all at once.

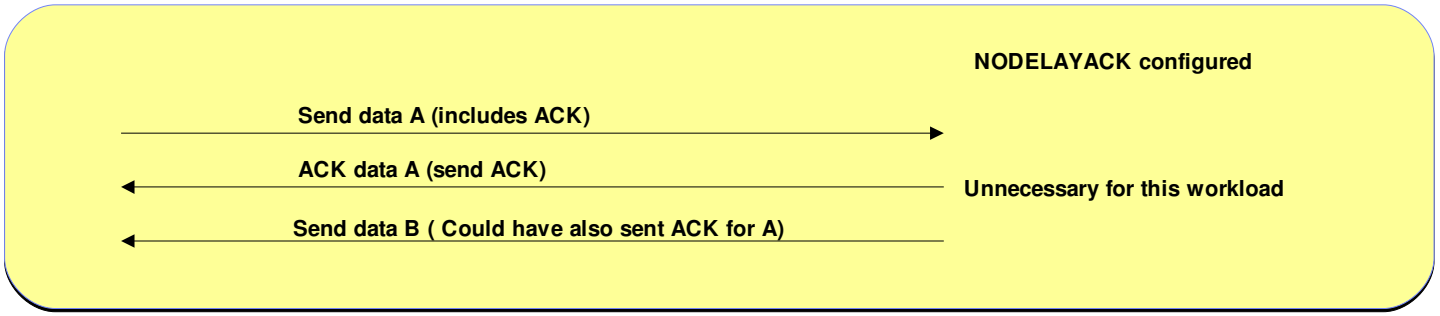
Goal: Minimize congestion

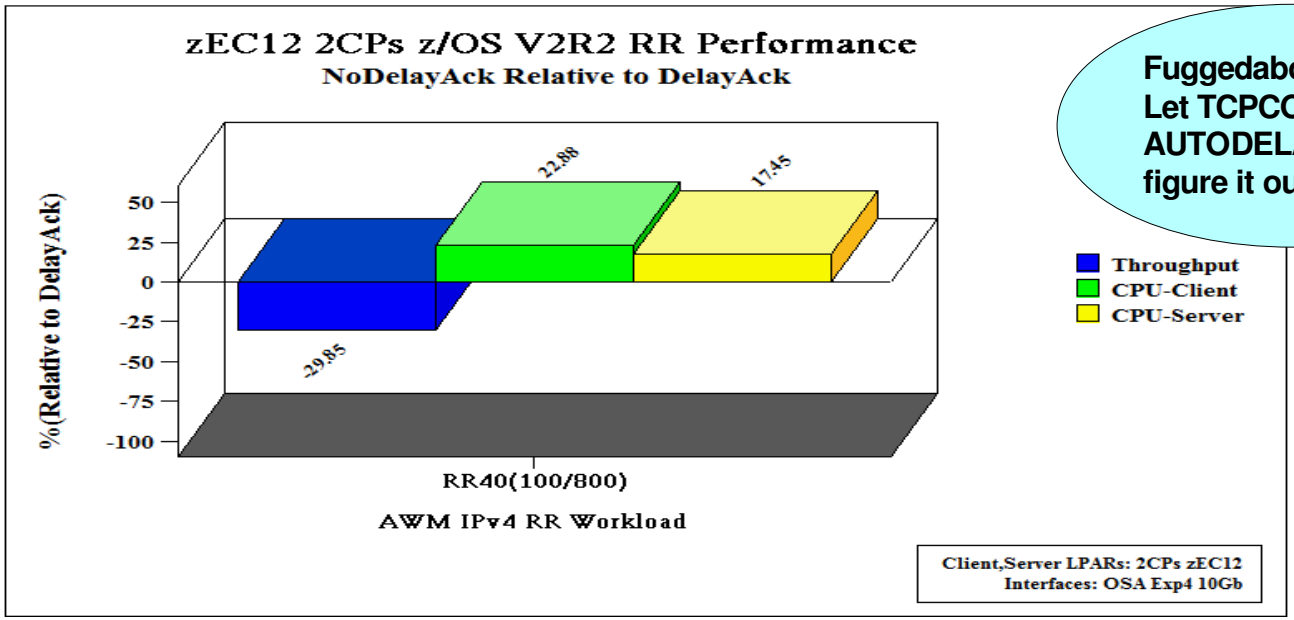
Delayed Acks (RFC1122)

A host that is receiving a stream of TCP data segments can increase efficiency in both the Internet and the hosts by sending fewer than one ACK segment per data segment received; this is known as a "Delayed ACK"

NODELAYACK ISSUE

NODELAYACK issue





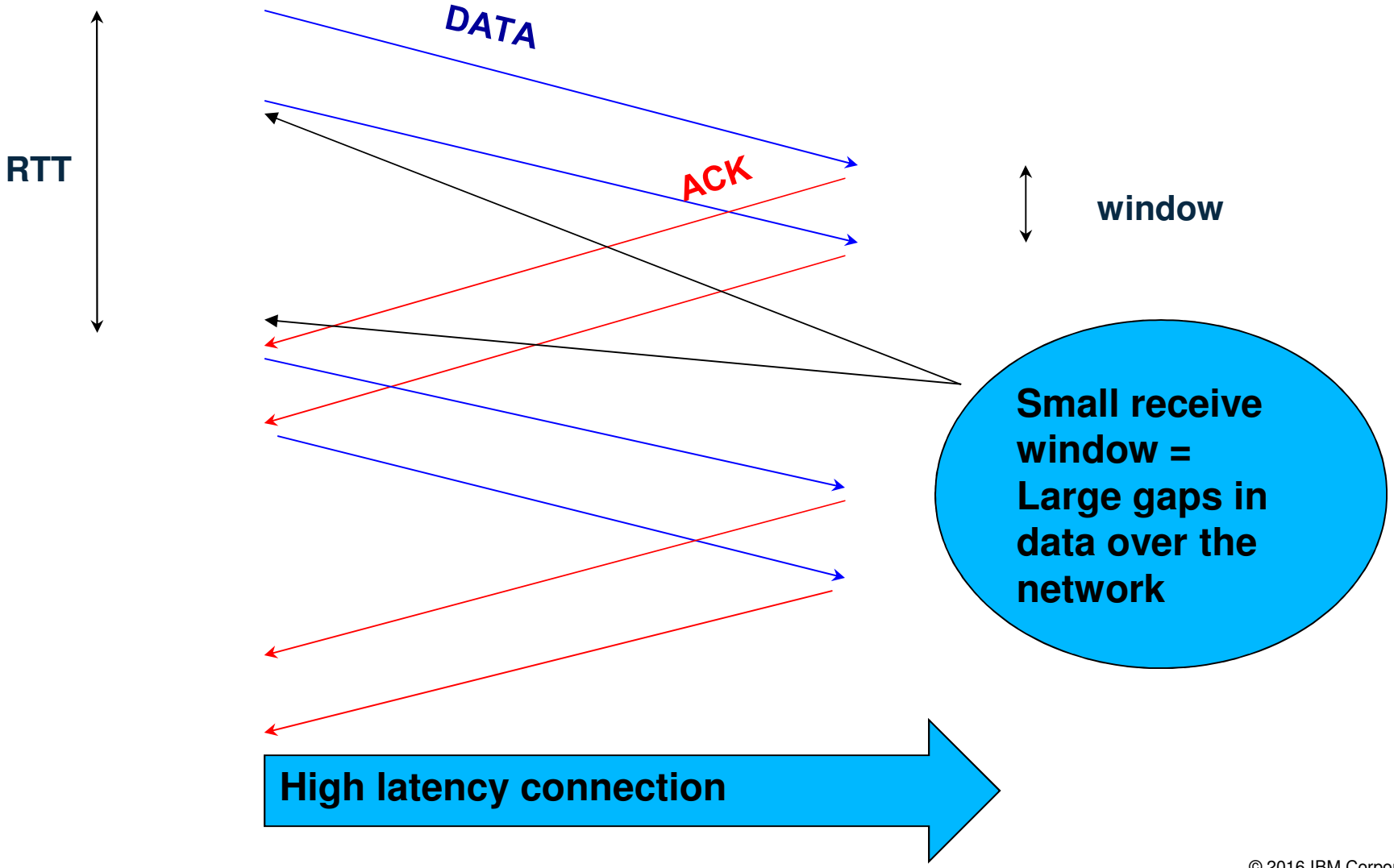
Fuggedaboutit!
Let TCPCONFIG AUTODELAYACK figure it out for you!

Dynamic Right Sizing

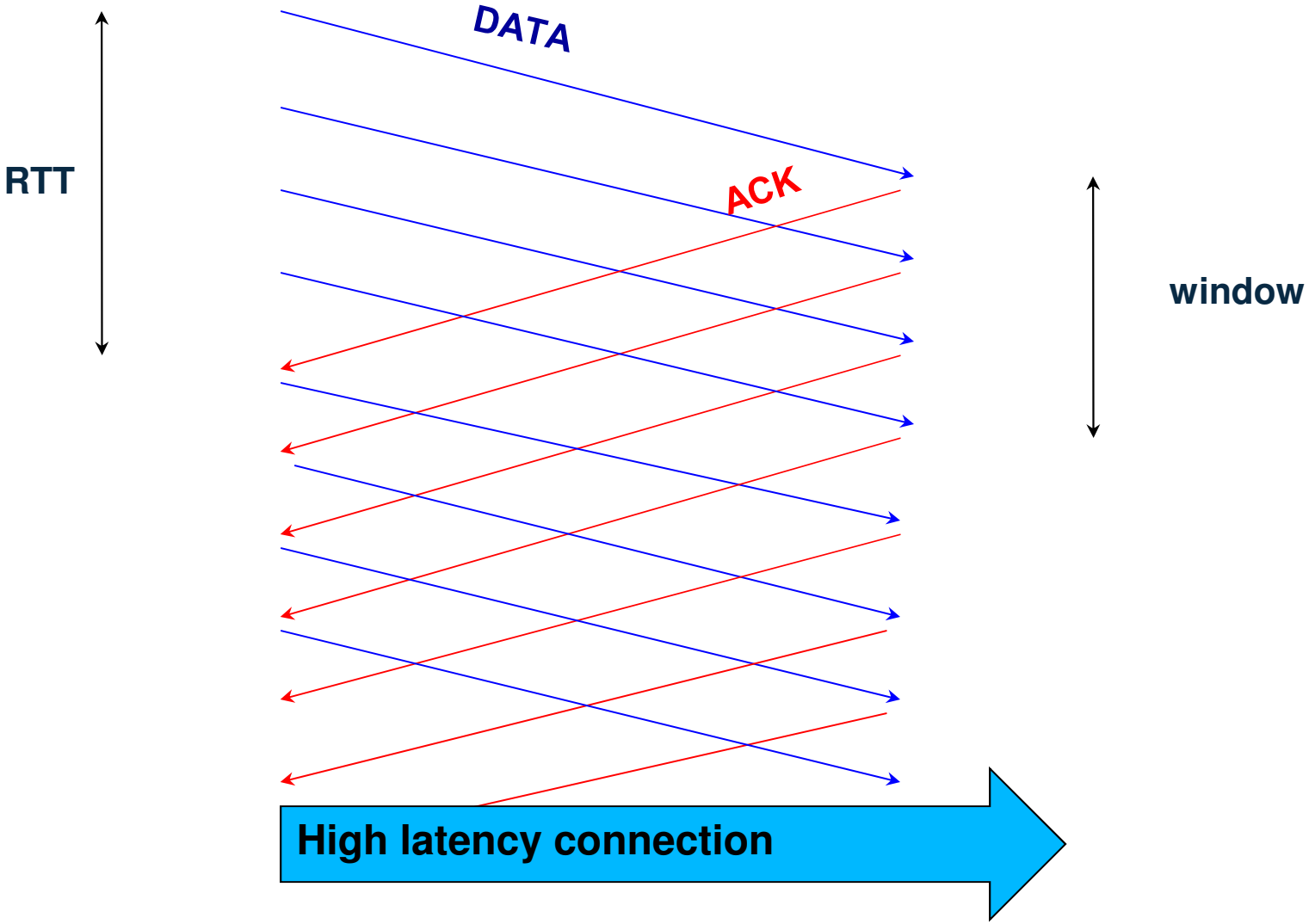
- Keep the pipe full and prevent sender from being constrained by the advertised window
- Improves performance for inbound streaming workloads over high latency networks with large bandwidth-delay product
- Function enabled automatically (no configuration)
- Has proven very helpful in several installations
- Stack dynamically increases the receive buffer size for the connection (in an attempt to not constrain the sender)
  - This in turn adjusts the advertised receive window
  - Allows window size to grow as high as 2M
- Allow enablement for connections that don't start as streamers
- Don't disable if not storage-constrained

Improved in V2R2!

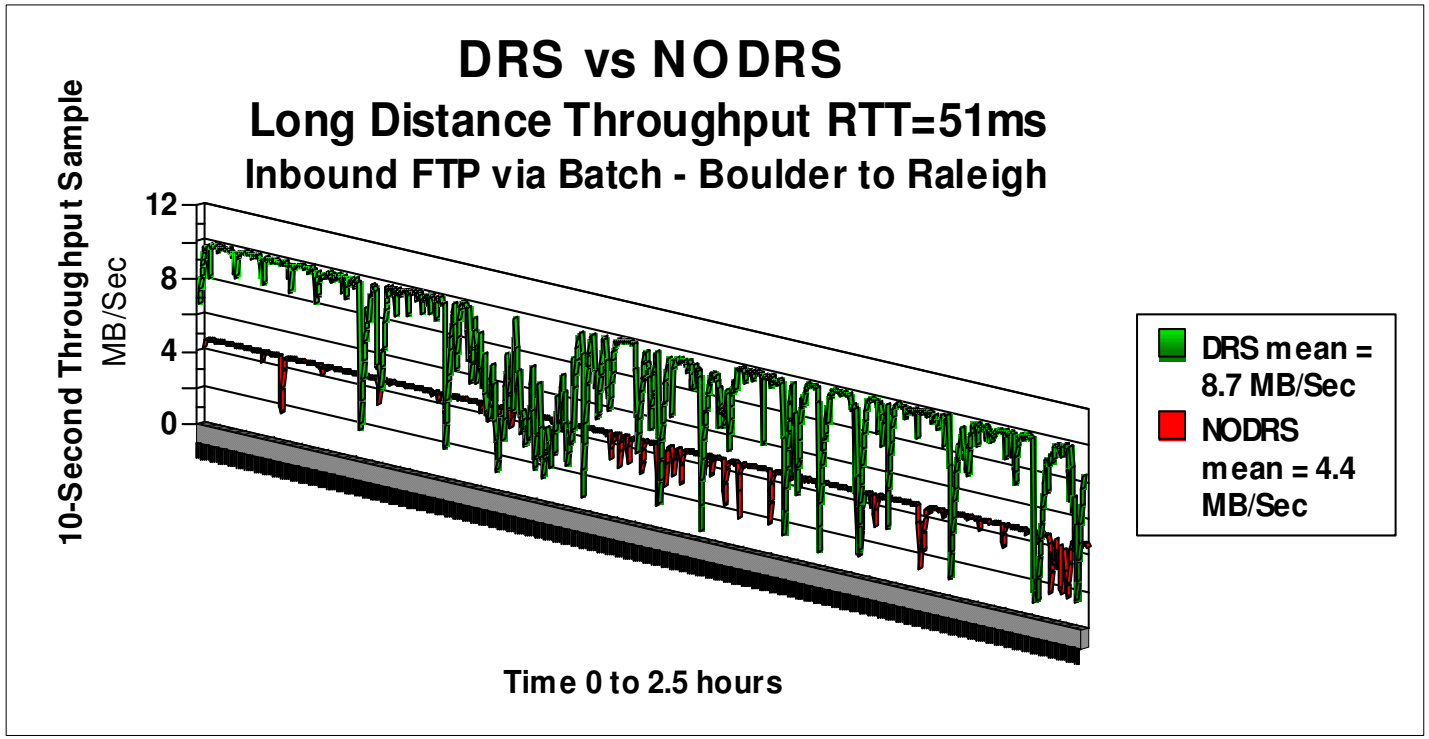
V2R2 Dynamic Right Sizing improvement



V2R2 Dynamic Right Sizing improvement



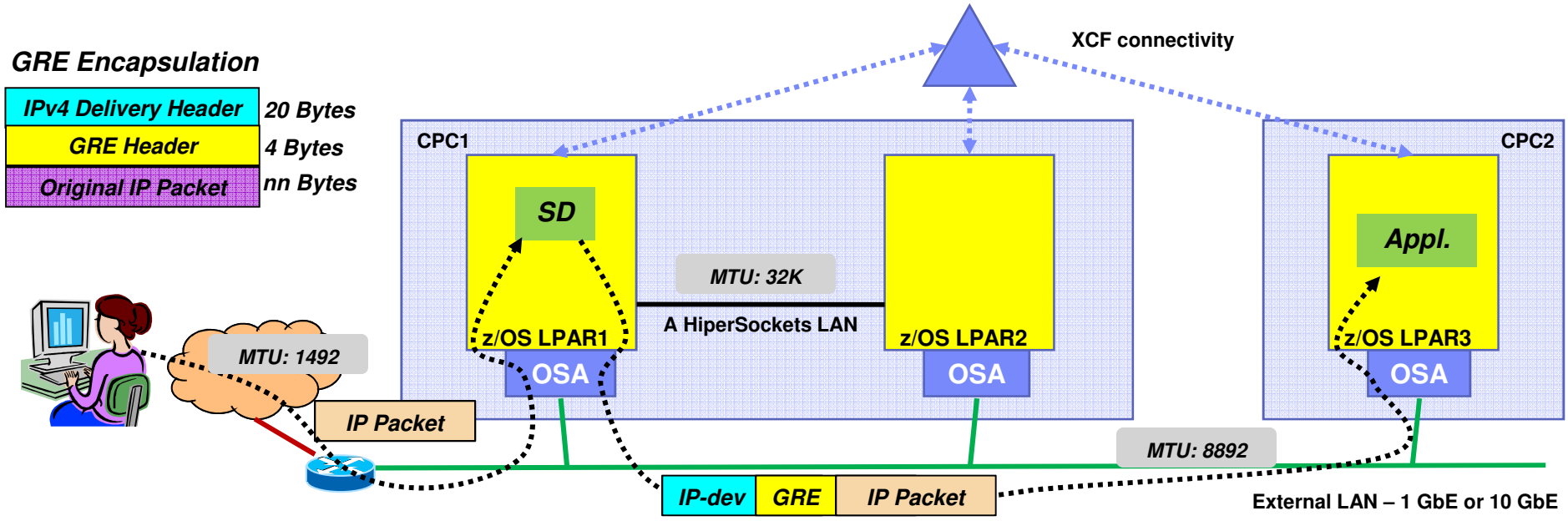
Dynamic Right Sizing



Over an extended 2.5 hour experiment, the DRS enabled receiver averaged double the throughput compared to no DRS.
This experiment repeatedly transferred a 2.8 GB file, and DRS never disabled over the 2.5 hour period.

VIPAROUTE and MTU size considerations

- When VIPAROUTE is used, the distributing stack adds a GRE header to the original IP packet before forwarding to the target stack
- Two ways to avoid fragmentation between distributing and target stacks:
  - Have clients use path MTU discovery
    - z/OS will factor in the GRE header size (24 bytes) when responding with next-hop MTU size
    - Not always possible to control distributed nodes' settings from the data center
  - Use jumbo-frames on the data center network
    - The access network will typically be limited to Ethernet MTU size (1492 bytes), while the data center network will be able to use jumbo frame MTU size (8892 bytes)
    - Adding the GRE header will not cause fragmentation in this scenario



VIPAROUTE fragmentation avoidance

- VIPAROUTE has been used extensively by many users to offload sysplex distributor forwarded traffic from XCF links
  - When used in combination with QDIO Accelerator for SD can result in dramatically reduced overhead for SD forwarding
- Fragmentation is still a concern for several customers
  - Resulting from the extra 24 bytes that are needed for the GRE header
  - Path MTU Discovery helps but doesn't solve the issue in some environments (where ICMP messages cannot flow across FWs)
  - Fragmentation can cause significant performance degradation

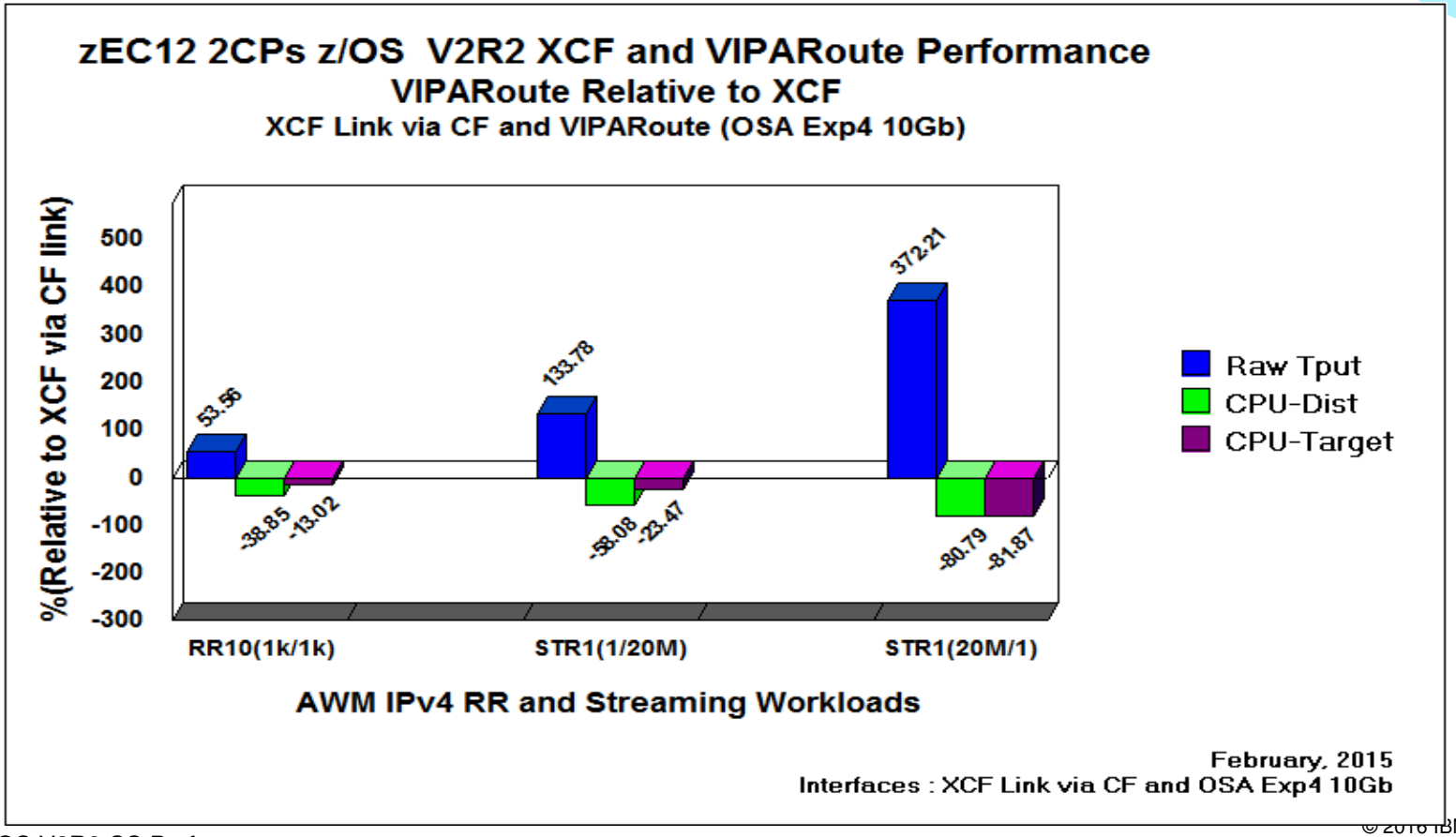
VIPAROUTE fragmentation avoidance (cont)

- V2R2 introduces a new autonomic option that will automatically reduce the MSS (Maximum Segment Size) of a distributed connection by the length of the GRE header.
  - This will allow the client TCP stack to build packets that allow for the 24 bytes of the GRE header to be added without any fragmentation being required.
  - ADJUSTDVIPAMSS on GLOBALCONFIG
    - Defaults to *AUTO* - Enables adjusted MSS
      - On target TCP/IP stacks when VIPAROUTE is being used
      - On Sysplex Distributor stack if it is also a target and VIPAROUTE is defined
    - Option *ALL* – Enables adjusted MSS for all connections using a DVIPA (distributed or not)
    - Option *NONE* - If you are already exploiting VIPAROUTE and know that there's no fragmentation possible in your environment you can disable this function
    - *Note:* This is a stack specific option. If the default is not taken then it must be configured on all systems (and TCP/IP stacks) in the sysplex
  - *VIPAROUTE fragmentation avoidance is also available via PTF on V1R13 (APAR PI38540) and V2R1 (APAR PI39519)*
  - This function is Off by default in the APARs

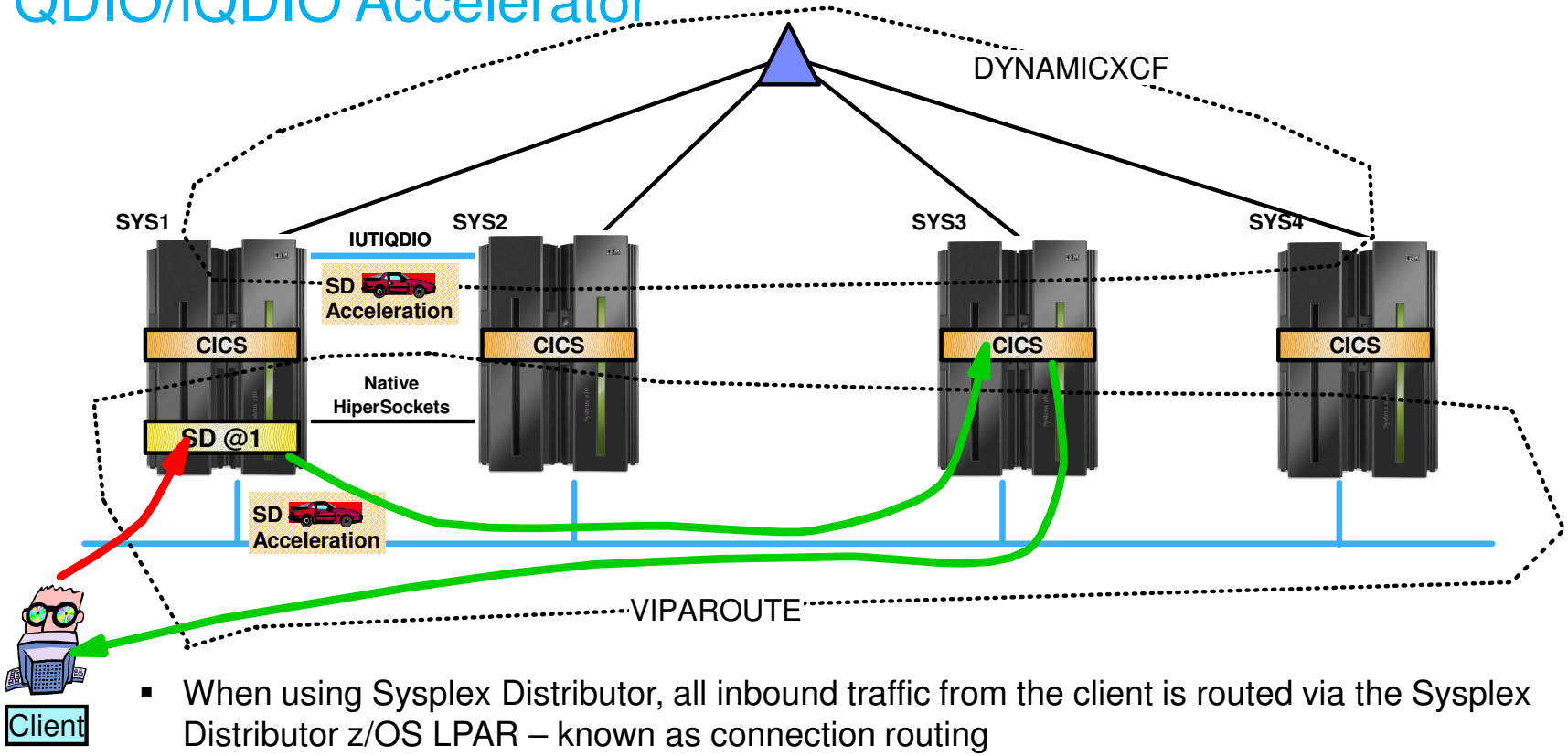
AUTOADJUSTMSS for VIPAROUTE: New GLOBALCONFIG parameter

- Automatically reduce the MSS (Maximum Segment Size) of a distributed connection by the length of the GRE header
  - Eliminate fragmentation
- Why you should be using VIPAROUTE for **Sysplex Distributor Workloads:**

Improved in V2R2!



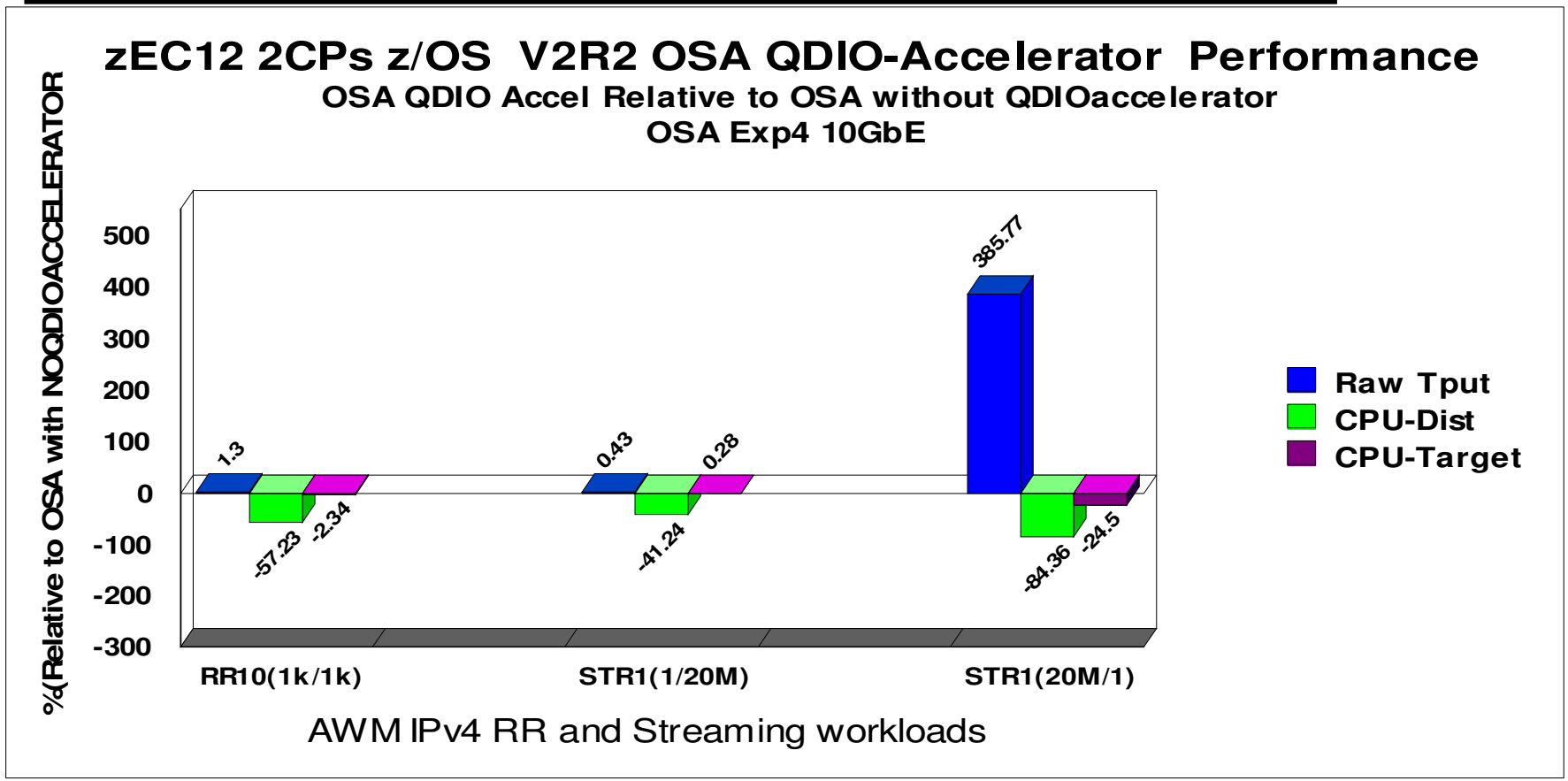
Sysplex Distributor connection routing benefits from QDIO/iQDIO Accelerator



- When using Sysplex Distributor, all inbound traffic from the client is routed via the Sysplex Distributor z/OS LPAR – known as connection routing
  - Outbound traffic goes directly back to the client
- When inbound packets to Sysplex Distributor is over QDIO or iQDIO (HiperSockets), Sysplex Distributor will perform accelerated connection routing when outbound is a DYNAMICXCF iQDIO interface - or when the outbound interface is a QDIO network interface
  - Helping reduce CPU overhead and latency in the Sysplex Distributor LPAR (SYS1)

Sysplex Distributor Accelerator Performance

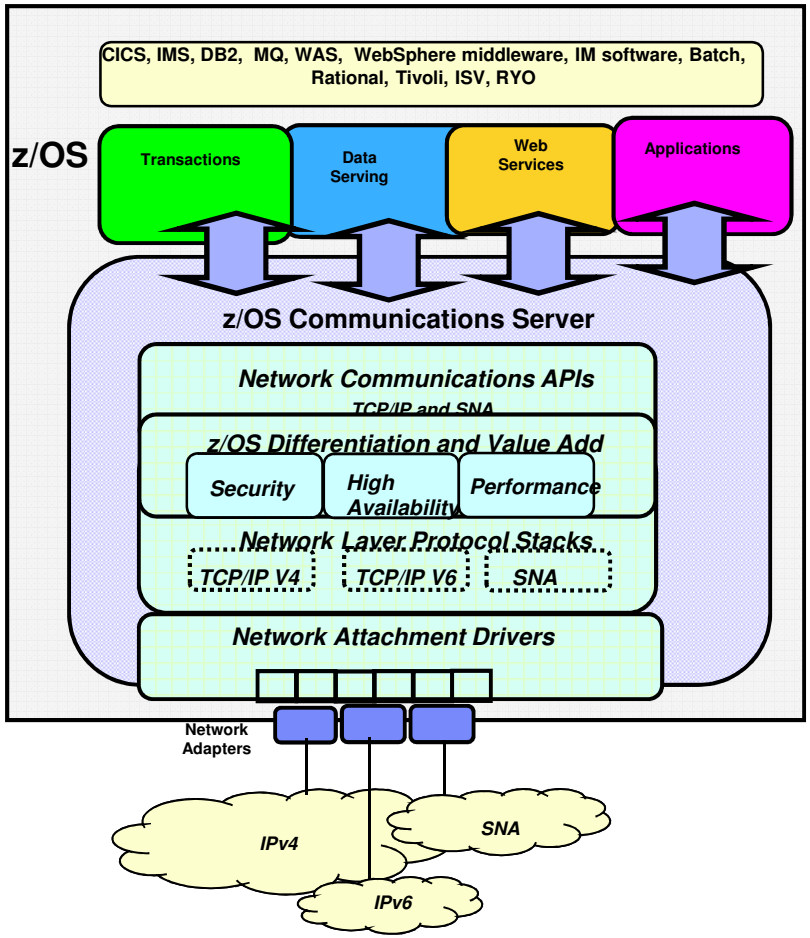
- ✓Intended to benefit all existing Sysplex Distributor users
- ✓Request/Response data pattern (1K request, 1K response, 10 concurrent sessions) – RR10
- ✓Streaming data pattern (1/20M – 1 byte in, 20MB response, 20M/1 – 20MB in, 1 byte response)
- ✓Percentages relative to no acceleration (both using VIPAROUTE and 10GbE OSA Express 4)



Note: The performance measurements discussed in this presentation are preliminary z/OS V2R2 Communications Server numbers and were collected using a dedicated system environment. The results obtained in other configurations or operating system environments may vary.

64-Bit enablement of the TCP/IP stack and DLCs

- TCP/IP has supported 64-bit applications since 64-bit support was introduced on the platform
  - But the mainline path has been 31-bit with extensive use of AR mode
- As systems become more powerful, customers have increased the workloads on the systems which in turn increases the storage demands placed on the systems.
- The storage in 31-bit addressing mode (below the bar) has been of special concern. Over the past several releases work has started to move storage that used to be obtained below the bar to 64-bit addressing mode (above the bar).
- Some of these changes, such as V1R13's move of the CTRACE and VIT above the bar, were visible to customers, while others were just changes in internal "plumbing".
- The next step is a large one: To move most of the remaining storage above the bar without incurring an unacceptable overhead in switching between AMODE(31) and AMODE(64) requires the complete 64-bit enablement of the TCP/IP stack and device drivers (DLCs) OSA QDIO and HiperSockets.



64 bit storage savings

Telnet run, large number of TCP connections

Address Space and Storage Type	31 - Bit V2R1 (KB)	64 - Bit V2R2 (KB)	% change from V2R1
Telnet ECSA	1,575	145	-91
TCP/IP ECSA	9,188	6,593	-28
TCP/IP Private	275,338	43,332	-84

Removing workload growth constraints from below-the-bar Private and Common storage:

ECSA savings:

- **Dynamic storage/data buffers moved**

Private savings:

- **Key connection control block moved**

- Based on benchmarks of modeled z/OS TCP sockets based workloads with telnet traffic patterns. The benefits any user will experience will vary.
- With the enablement of 64 bit there is some performance impact on network CPU cost

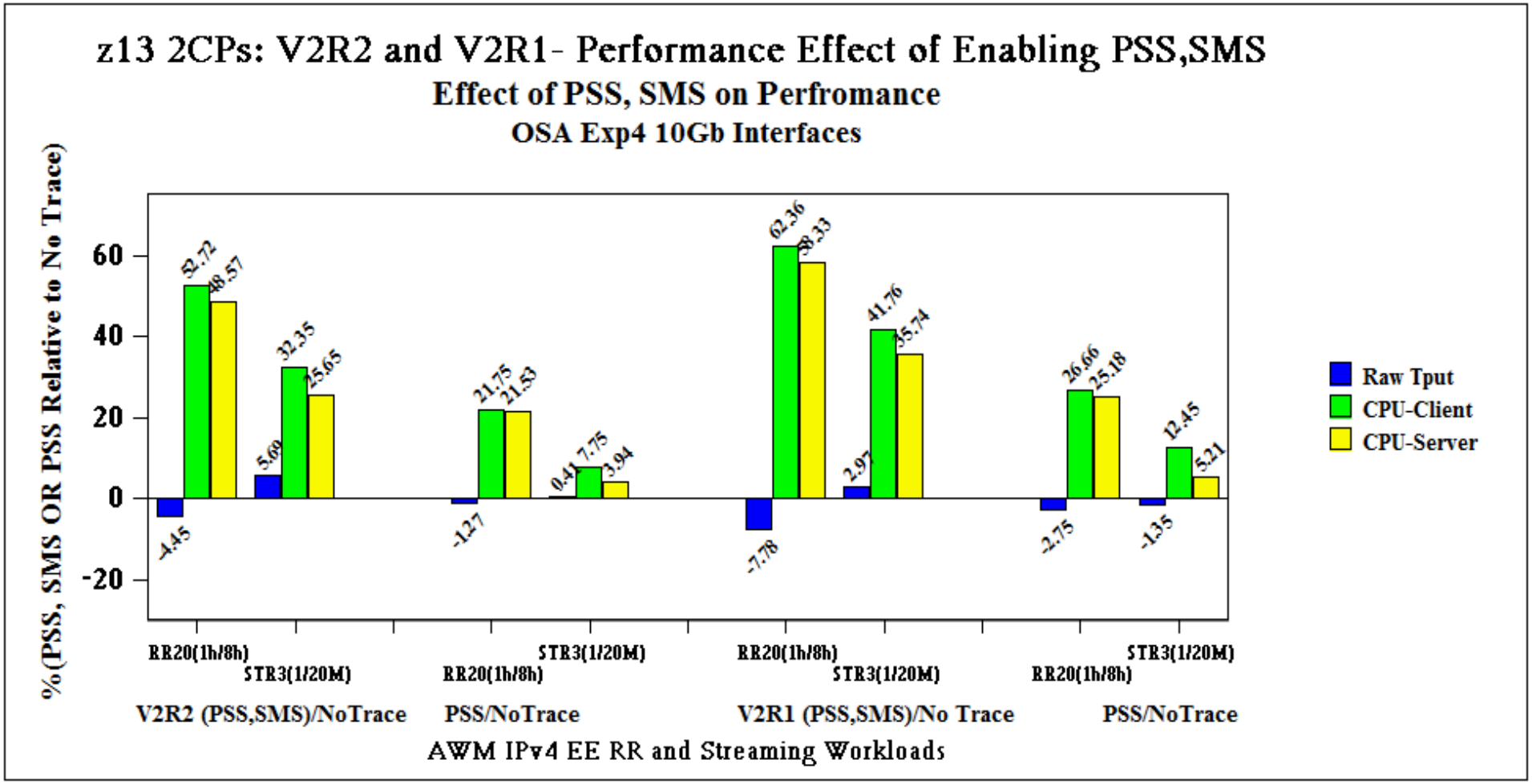
VTAM Internal Trace Improvements

- The VTAM Internal Trace (VIT) can be a valuable diagnostic tool used by IBM Level 2 to diagnose customer problems.
- Occasionally, during periods of high utilization, some VIT records would get “lost” before they could be captured to VIT storage.
- In V2R2, VTAM Internal Trace processing is enhanced to greatly reduce the likelihood of lost VIT records during high-volume activity.

VTAM Internal Trace - Recommendation

- There are eight VIT options that are enabled by default
  - API,CIO,MSG,NRM,PIU,PSS,SMS,SSCP
- Given the infrequent need for the SMS option during problem diagnosis, it is often not worth the CPU cost of the SMS option for the slight improvement in first failure data capture.
- Therefore, we believe that disabling the SMS VIT option is the best choice for most customers except those actively working to gather problem documentation under the direction of IBM Level 2 support.
- After VTAM is up, disable the SMS VIT option by entering the command:
  - MODIFY NET,NOTRACE,TYPE=VTAM,OPT=SMS
- APAR OA49999 will change the default option set to no longer include SMS
  - Projected availability in 2Q16
  - This APAR will also update the CSVTAM_VIT_OPT_PSSSMS health check to no longer warn if the SMS VIT option is not active (will become CSVTAM_VIT_OPT_PSS)

VTAM Internal Trace – Effect of PSS,SMS on Performance



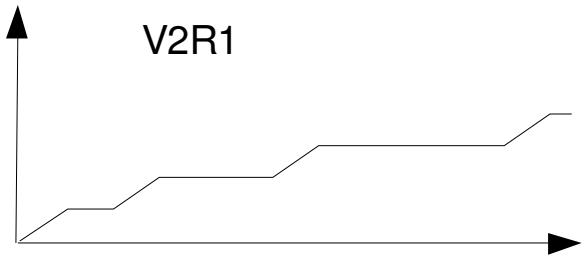
- As shown above, enabling both PSS and SMS compared to enabling only PSS causes significant impact on RR and streaming throughput and CPU cost. Therefore it is strongly recommended to disable SMS trace.

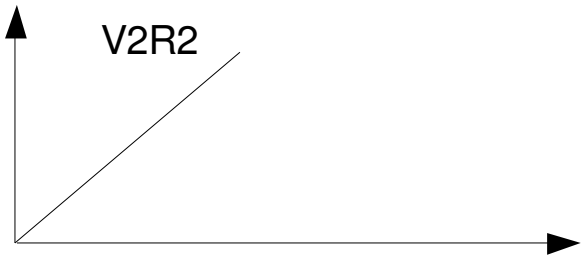
Enhanced IKED scalability (IPSEC)

- In V2R2, z/OS IKED is modified to handle heavy bursts of negotiations from very large numbers (multiple thousands) of IKE peers
  - –A new thread pool is added to parallelize handling of IKE messages from different peers
  - –Logic is added to minimize the amount of effort IKED spends processing retransmitted messages from peers
  - –Transparent to the vast majority of current IKED users
    - Improvement will be most noticeable to users with very large numbers (multiple thousands) of IKE peers

Enhanced IKED scalability

- z/OS V2R2 performance results show significant performance improvements in establishing SAs when a large number of concurrent client requests arrive in a small interval of time
  - IKE V1 (more messages exchanged):
    - Up to 6.8X improvement in throughput (rate of SA activations) and up to 75% reduction in CPU cost *
  - IKE V2:
    - Up to 3.8X improvement in throughput (rate of SA activations) and up to 57% reduction in CPU cost *

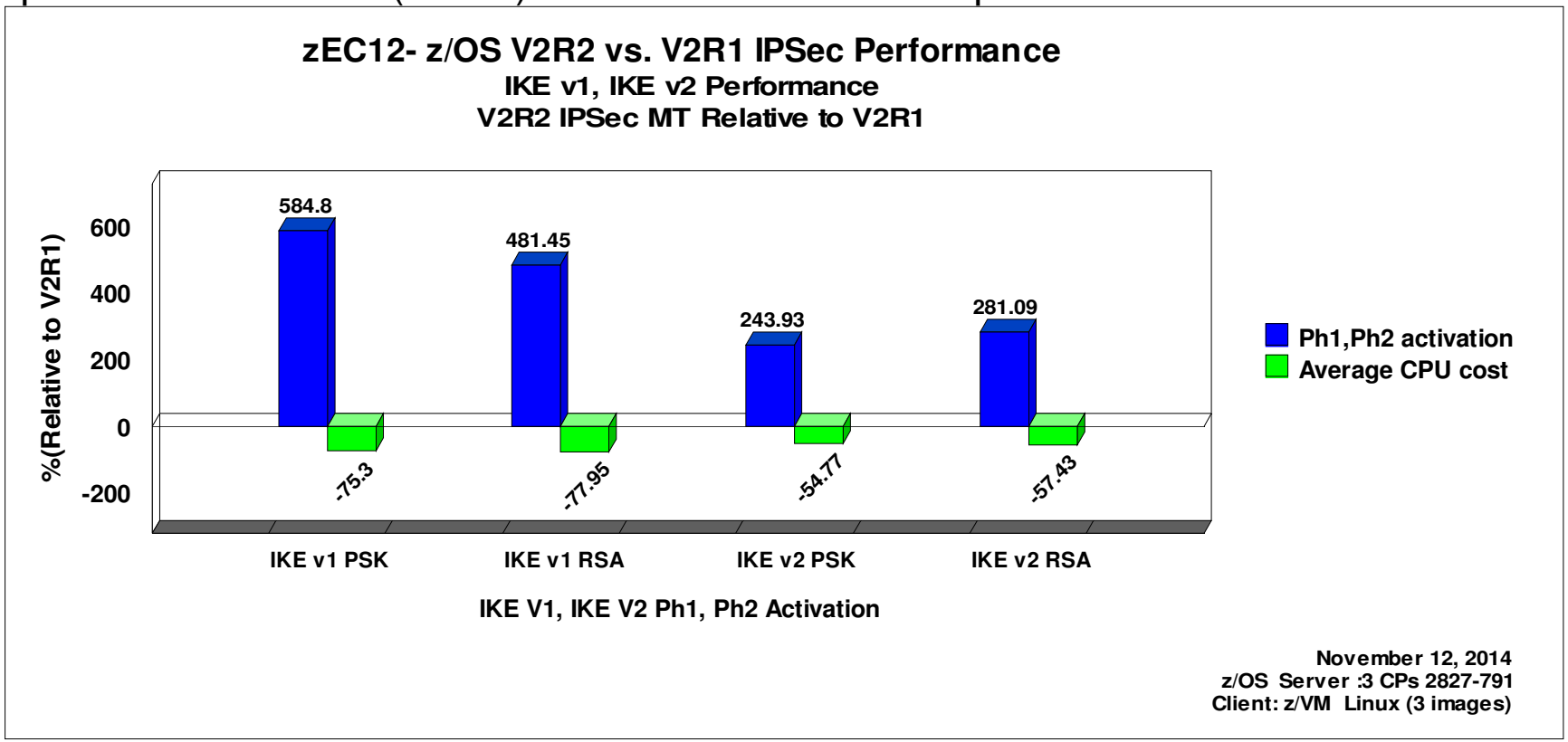




*Note: The performance measurements were collected in IBM internal tests using a dedicated system environment. 4,200 clients simulated using 4 Linux for System z images running under zVM. IKE v1 benchmarks performed with PSK. IKE v2 benchmarks performed with RSA. The results obtained in other configurations or operating system environments may vary significantly depending upon environments used. Therefore, no assurance can be given, and there is no guarantee that an individual user will achieve performance or throughput improvements equivalent to the results stated here.

V2R2 and V2R1 IKE v1,v2 Performance Summary

- Performance Comparison (V2R2 vs. V2R1)
- For IKE v1 PSK and RSA, V2R2 provides (481-584)% higher rate for ph1, ph2 activation with (75-78)% lower CPU cost compared to V2R1
- For IKE v2 PSK and RSA, V2R2 provides (244-281)% higher rate for ph1, ph2 activation with (55-57)% lower CPU cost compared to V2R1



V2R2 and V2R1 IKE v1,v2 Performance Summary ..

- Performance Comparison (V2R2 vs. V2R1)
- V2R1 and V2R2 IKED v1,V2 Average SA activation rate and CPU cost per activation

Description	IKED v1		IKED v2	
	trans/sec	CPU Cost	trans/sec	CPU Cost
		PSK		
V2R1	6.69	179.42	10.60	115.99
V2R2	45.80	44.31	36.47	52.46
		RSA		
V2R1	5.19	232.28	7.15	166.64
V2R2	30.19	51.12	27.23	70.94

- **Note: Here trans/sec provides rate of ph1, ph2 activation**
- **CPU Cost is the average cost per activation and it is indicated in msec unit on zEC12**

SMC Applicability Tool

Evaluating SMC applicability and benefits SMC Applicability Tool (SMCAT)

As customers express interest in SMC-R/SMC-D one of the initial questions asked is:

- "What benefit will SMC provide in my environment?"
  - Some users are well aware of significant traffic patterns that can benefit from SMC
  - But others are unsure of how much of their TCP traffic (in their environment) is:
    - z/OS to z/OS
    - IPSEC?
    - Traffic well suited to SMC?

- Reviewing SMF records, using Netstat displays, Ctrace analysis and reports from various Network Management products can provide these insights…

This approach can be a time consuming activity that requires significant expertise.

SMC Applicability Tool Introduction

A new tool called SMC Applicability Tool (SMCAT) has been created that will help customers determine the ***potential*** value of SMC in their environment with minimal effort and minimal impact

- SMCAT is integrated within the TCP/IP stack:
  Gather new statistics that are used to project SMC applicability and benefits for the current system
  - Minimal system overhead, no changes in TCP/IP network flows
  - Produces reports on potential benefits of enabling SMC
- Available via the service stream on existing z/OS releases as well
  - V1R13 PI48309 / UI31050
  - V2R1 PI48155 / UI31054
  - V2R2 PI48155 / UI31055

  Does not require:
  - SMC code or RoCE hardware to use
  - Any changes in IP configuration (i.e. captures your normal TCP/IP workloads)

SMCAT Usage Overview

Activated by Operator command
(***Vary TCPIP,,SMCAT,dsn(smcatconfig)*** – Input dataset contains:

- Interval Duration, list of IP addresses or IP subnets of peer z/OS systems (i.e. systems that we can use SMC for)
  - If subnets are used, the entire subnet must be comprised of z/OS systems that are SMC eligible
  - It is important that all the IP addresses used for establishing TCP connections are specified (including DVIPAs, etc.)
- At the end of the interval a summary report is generated that includes:
  1. ***Percent of traffic eligible for SMC*** *(% of TCP traffic that is eligible for SMC)*
     - *All traffic that matches configured IP addresses (not using IPSec or FRCA)*
  2. ***Percent of traffic well suited for SMC*** *(your eligible traffic that is also "well suited" to SMC, excludes workloads with very short lived TCP connections that have trivial payloads)*
     - *Includes break out of application send and recv sizes (bigger is better!)*
     - *Helps users quantify SMC benefit (reduced latency / reduced CPU cost)*

SMCAT Usage Overview (continued)

The Summary Report includes 2 sections based on the specified IP addresses/subnets defined in SMCAT configuration file:

1. Potential benefit:

   All TCP traffic that matches the configuration - Includes TCP traffic that could not use SMC without changes (TCP traffic that does not meet the direct IP route connectivity requirement)

   This represents the opportunity of re-configuring routing topology to enable SMC

1. Immediate benefit:

   The TCP traffic that can use SMC immediately / as is (meets SMC direct route connectivity requirements). Subset of section 1.

   Detected by the tool automatically (non-routed traffic)

SMC Applicability Tool Sample Report (Direct Connections)

```
Interval Details:
    Total TCP Connections:                                      100
    Total SMC eligible connections:                              15
          Total SMC well-suited connections:                     14
    Total outbound traffic (in segments)                       1000
          SMC well-suited outbound traffic (in segments)        150
    Total inbound traffic (in segments)                         500
          SMC well-suited inbound traffic (in segments)          70

    Application send sizes used for well-suited connections:
      Size                                  # sends     Percentage
      ----                                  -------     ----------
      1500 (<=1500):                             15      37%
      4K (>1500 and <=4k):                        7      17%
      8K (>4k and <= 8k):                         3       7%
      16K (>8k and <= 16k):                       4      10%
      32K (>16k and <= 32k):                      8      20%
      64K (>32k and <= 64k):                      3       7%
      256K (>64K and <= 256K):                    1       2%
      >256K:                                      0       0%

    Application receive sizes used for well-suited connections:
      Size                                  # recvs     Percentage
      ----                                  -------     ----------
      1500 (<=1500):                              8      38%
      4K (>1500 and <=4k):                        3      14%
      8K (>4k and <= 8k):                         2      10%
      16K (>8k and <= 16k):                       2      10%
      32K (>16k and <= 32k):                      4      20%
      64K (>32k and <= 64k):                      1       5%
      256K (>64K and <= 256K):                    1       5%
      >256K:                                      0       0%

    -------------------SMCAT Summary Report Export Area------------------

    20,10,4,5,10,5,2,0
    10,5,3,3,5,2,2,0
    15,7,3,4,8,3,1,0
    8,3,2,2,4,1,1,0

    -------------------End Export Area-----------------------------------
```

How much of my TCP workload can benefit from SMC?

What kind of CPU savings can I expect from SMC?

This is all of the send and receive data provided in a new export area (Send to IBM).

Configuring the SMCAT Dataset

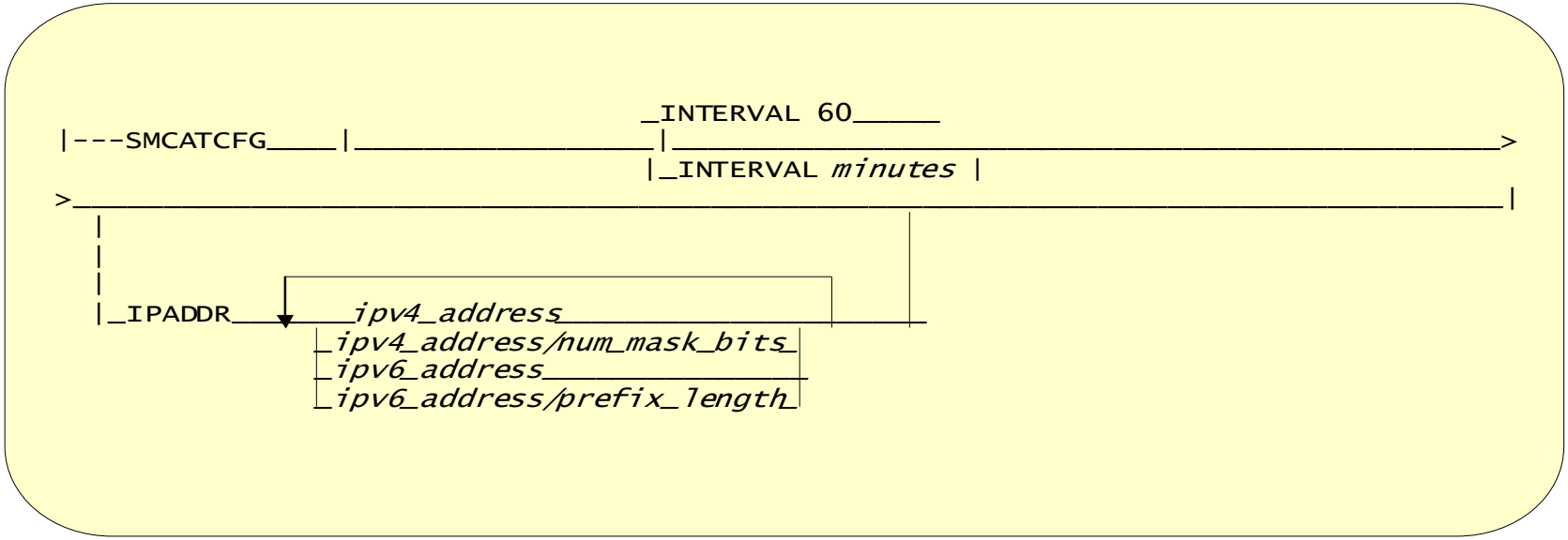
SMCAT data set configuration

- Interval defaults to 60 minutes

Max interval is 1440 minutes (24 hours)

- IPADDR is a list of IPv4 and Ipv6 addresses and subnets

256 max combination of addresses and subnets



SMCAT Dataset Example

```
SMCATCFG INTERVAL 120
IPADDR
C5::1:2:3:4/126
9.67.113.61
```

Simple!

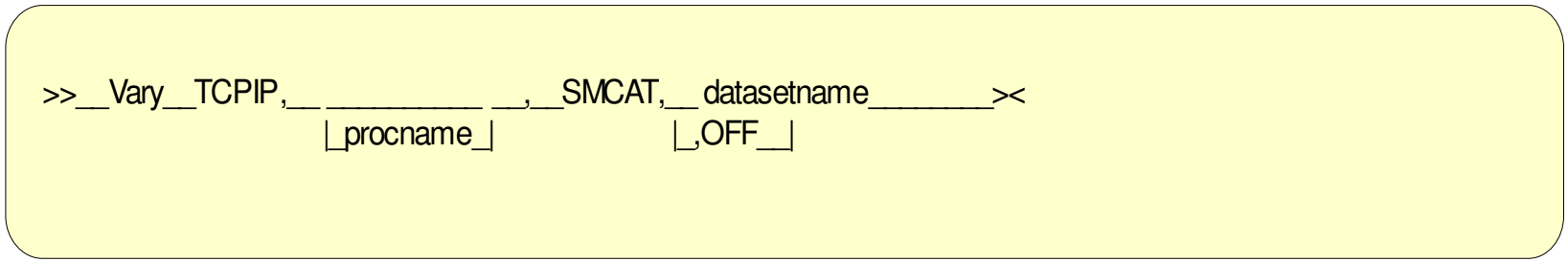
When SMCAT is started using this SMCAT configuration data set it will:

- **Monitor TCP traffic for 2 hours for:**
  - **IPv6 prefix C5::1:2:3:4/126 and**
  - **IPv4 address 9.67.113.61**

Starting and Stopping SMCAT

Vary TCPIP,,SMCAT command starts and stops the monitoring tool:

- **datasetname value indicates that SMCAT is being turned on**
- **datasetname contains the SMCATCFG statement that specifies monitoring interval and IP addresses or subnets to be monitored**
- **OFF will stop SMCAT monitoring and generate report**




```
VARY TCPIP,TCPPROC,SMCAT,USER99.TCPIP.SMCAT1
```

SMCAT Usage Notes:

- When you have many instances of hosts that provide similar workloads (similar application servers) consider measuring a subset of the hosts and then extrapolating the SMCAT results of your sample across your enterprise data center
- Run the SMCAT tool at different intervals to measure changing workloads

Reference Information

z/OS CS Performance References

- **z/OS Communications Server performance index:**
  This is an index to all published performance information for the z/OS Communications Server. This index is updated when updates are made to existing documentation or additional documentation is added. You may want to bookmark this link.

  http://www.ibm.com/support/docview.wss?rs=852&uid=swg27005524

- **SHARE presentations (http://www.share.org)**
  Share 2016 Winter Technical conference (San Antonio)

  - **z/OS V2R2 Communications Server Technical Update, Part I and II (sessions 18517 ad 18518)**
  - **Sysplex Networking Technologies and Considerations (session 18521)**
  - **z/OS Communications Server Performance Functions Update (session 18526)**

Additional Information

URL	Content
http://www.ibm.com/systems/z	IBM Enterprise Servers (zSeries & S/390)
http://www.ibm.com/systems/z/hardware/networking	zSeries Networking
http://www.ibm.com/software/products/us/en/commserver	IBM Communications Severs
http://www.ibm.com/software/products/us/en/commserver-zos	z/OS Communications Server
http://www.ibm.com/software/network/commserver/zos/support	z/OS Communications Server Technical Support
http://www.facebook.com/IBMCommserver	z/OS Communications Server facebook page
http://twitter.com/IBM_Commserver	z/OS Communications Server on twitter
http://www.ibm.com/systems/z/os/zos/bkserv/v2r2pdf	z/OS Communications Server product library
http://www.redbooks.ibm.com	ITSO Redbooks
http://www.ibm.com/support/techdocs	Technical Information Data Base (Flashes, Presentations, White Papers, etc.)
http://www.ibm.com/software/products/awm	IBM Application Workload Modeler (AWM)
http://www.ibm.com/software/products/workloadsimulator	IBM Workload Simulator (IWS; aka TPNS)
http://www.ibm.com/support/docview.wss?rs=852&uid=swg27005524	z/OS Communications Server Performance