z/OS V2R3 Communications Server Performance Summary

Christopher Nyamful
cnyamfu@us.ibm.com
Dan Patel
danpatel@us.ibm.com
Dave Herr
dherr@us.ibm.com

Contents

Contents (cont'd)

2.3 Network Security Deployment on z/OS

Contents (cont’d)

Contents (cont’d)

Contents (cont’d)

Trademarks and notices

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

BigInsights
BlueMix
CICS*
COGNOS*
DB2*
DFSMSdfp
DFSMSdss
DFSMShsm
DFSORT
DS6000*
DS8000*
FICON*
GDPS*
HyperSwap
IBM*
IBM (logo)*
IMS
Language Environment*
MQSeries*
Parallel Sysplex*
PartnerWorld*
RACF*
Rational*
Redbooks*
REXX
SmartCloud*
System z10*
Tivoli*
UrbanCode
WebSphere*
z13
z14
zEnterprise*
z/OS*
zSecure
z Systems
z/VM*

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

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 therefrom.
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.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.
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.
Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and
Linux is a registered trademark of Linus Torvalds 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.
OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.
TEALEAF is a registered trademark of Tealeaf, an IBM Company.
Windows Server and the Windows logo are trademarks of the Microsoft group of countries.
Worklight is a trademark or registered trademark of Worklight, an IBM Company.
UNIX is a registered trademark of The Open Group in the United States and other countries.
* Other product and service names might be trademarks of IBM or other companies.

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.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g, zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the “Authorized Use Table for IBM Machines” provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html (“AUT”). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Performance Disclaimers

- The performance measurements discussed in this document were collected using a dedicated system environment. 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. Users of this document should verify the applicable data for their specific environment.

- The CPU numbers listed in the presentation includes all z/OS host networking related CPU overhead (including dispatching costs) from the network device driver layer up through the application socket layer. The socket applications used in the micro-benchmarks for this publication have no application logic, so the CPU numbers represent the total application cost which in this case equals the network related costs. In real workloads, networking related CPU cost is a small fraction of the overall application transaction cost.

V2R3 NETWORK PERFORMANCE

- ✓ Inbound Workload Queue Performance

V2R3 NETWORK PERFORMANCE

1.1 Inbound Workload Queue Performance

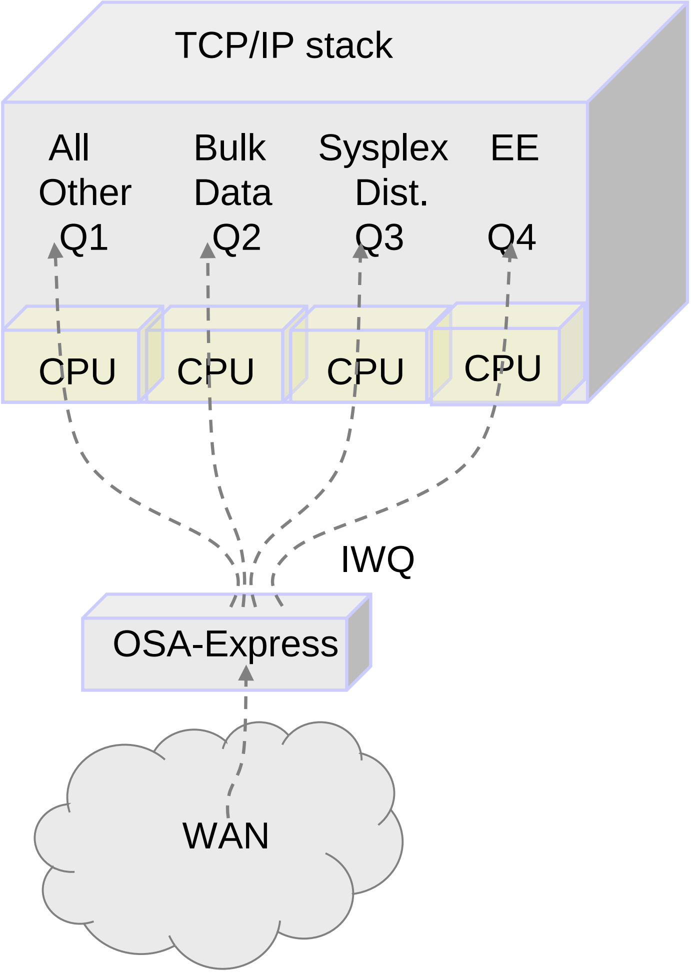
1.1a IWQ Performance – Background

- OSA separates inbound packets and routes them over four different ancillary input queues on the same interface

- Bulk data (such as FTP)
- Sysplex Distributor (SD)
- Enterprise Extender (EE)
- All other traffic (primary)

- z/OS can service each queue concurrently using separate processors (this implies separate CPU or separate worker threads)

- Stack receives pre-sorted packets



V2R3 NETWORK PERFORMANCE…

1.1 Inbound Workload Queue Performance

1.1b Workload definitions for performance measurements

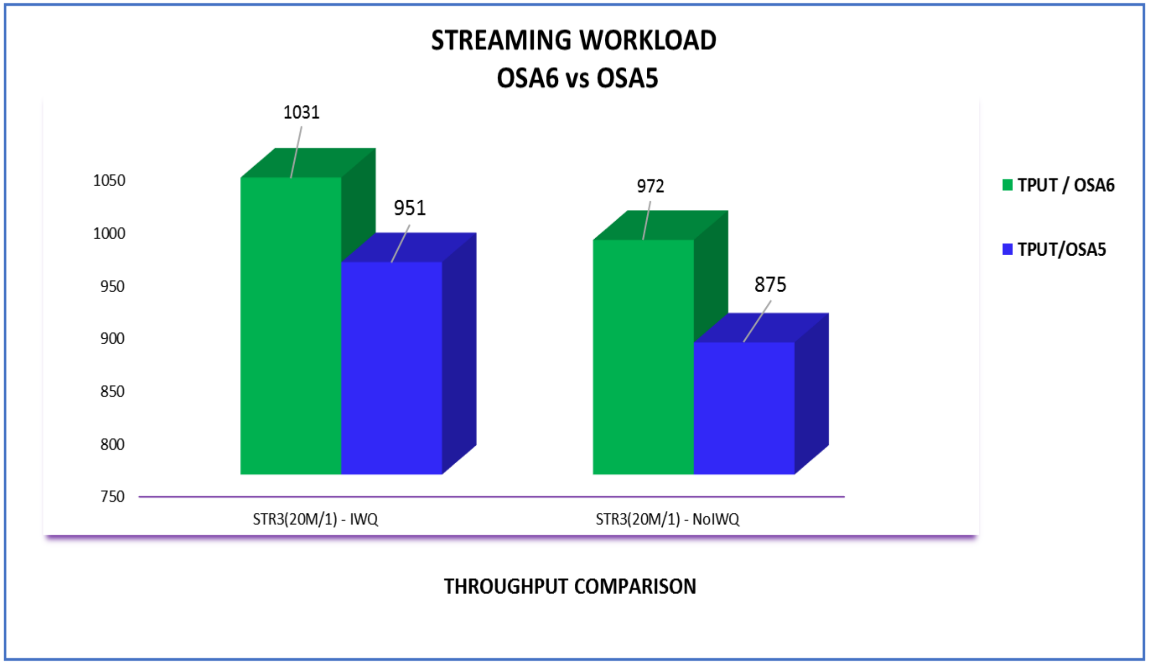
- RR10 1K/1K – 10 TCP request response connections sending and receiving 1000 bytes (persistent connections)
- STR3 20M/1 – 3 TCP streaming connections sending 20,000,000 bytes and receiving 1 byte
- All results collected on z14 running z/OS V2R3
- Results obtained using very lightweight applications – no application logic
- 2 CPs per LPAR
- Dedicated OSAs (OSA Exp6S 10Gb and OSA Exp5s 10Gb)
- Tests with and without enabling IWQ
  - IWQ enabled with (INTERFACE configuration Option ‘INBPERF DYNAMIC WORKLOADQ’)

V2R3 NETWORK PERFORMANCE…

1.1 IWQ Performance (OSA Exp6s vs. OSA Exp5s)

1.1c Streaming Workload Performance –Clear Text

- OSA Exp6s vs OSA Exp5s – Streaming workload (send 20M / receive 1)
  - OSA Exp6s improved throughput for streaming workload for both IWQ and non-IWQ compared to OSA Exp5s.
  - OSA Exp6s with IWQ increases throughput by 8.3% compare to OSA Exp5s with IWQ enabled
  - OSA Exp6s without IWQ improves throughput by 11% versus OSA Exp5s without IWQ
  - Note: IWQ provides higher throughput (eliminates out-of-order



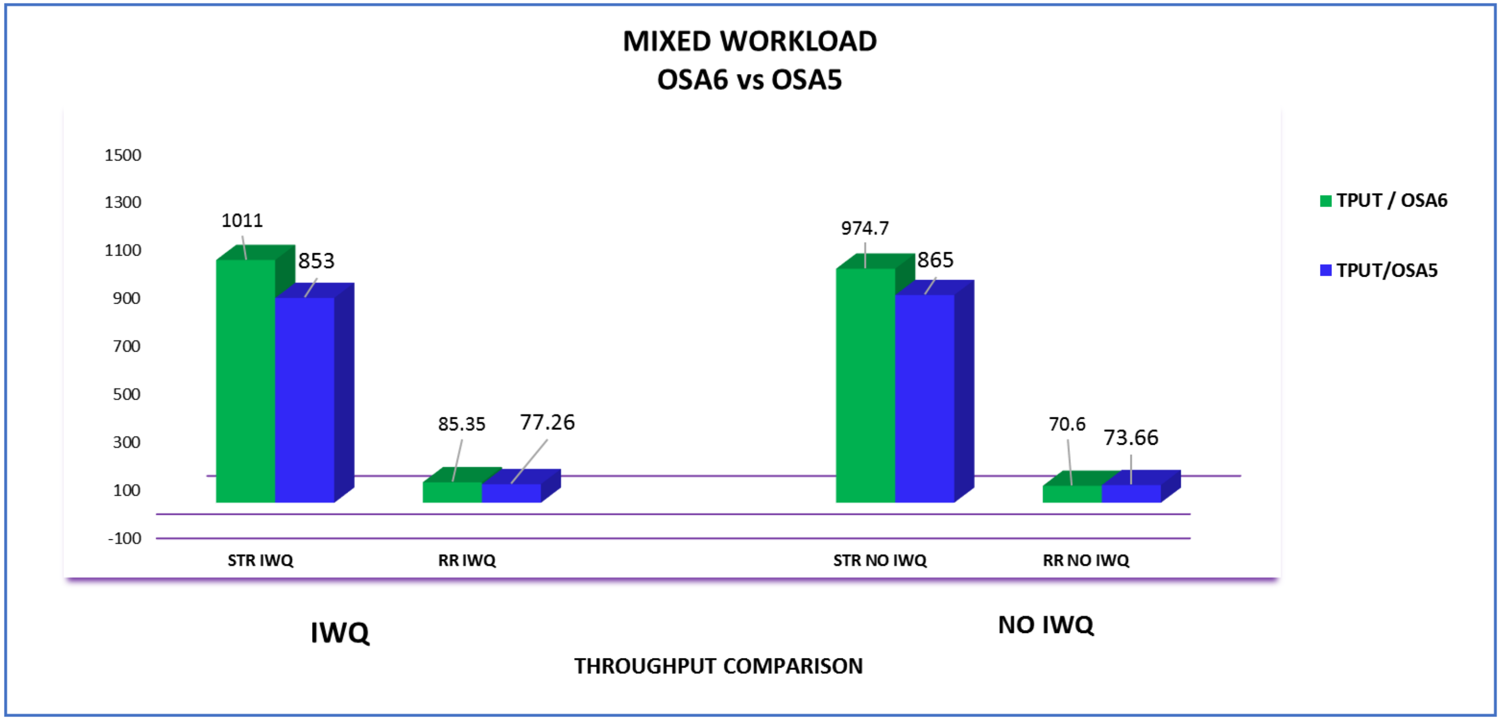
- Note: Throughput for three streaming sessions is represented in MB/Sec

V2R3 NETWORK PERFORMANCE…

1.1 IWQ Performance (OSA Exp6 vs OSA Exp5s) ….

1.1d Mixed Workloads Performance – Clear Text

- OSA Exp6s vs OSA Exp5s – Mixed workloads (RR and STR) with and without IWQ
  - OSA Exp6s improves throughput for streaming workload (with or without IWQ)
  - Request Response workload with OSA Exp6s IWQ shows 20% throughput increase vs NO IWQ (85 vs 70) MB/Sec



Note: Throughput for streaming and request response workloads are represented in MB/Sec

V2R3 NETWORK SECURITY PERFORMANCE

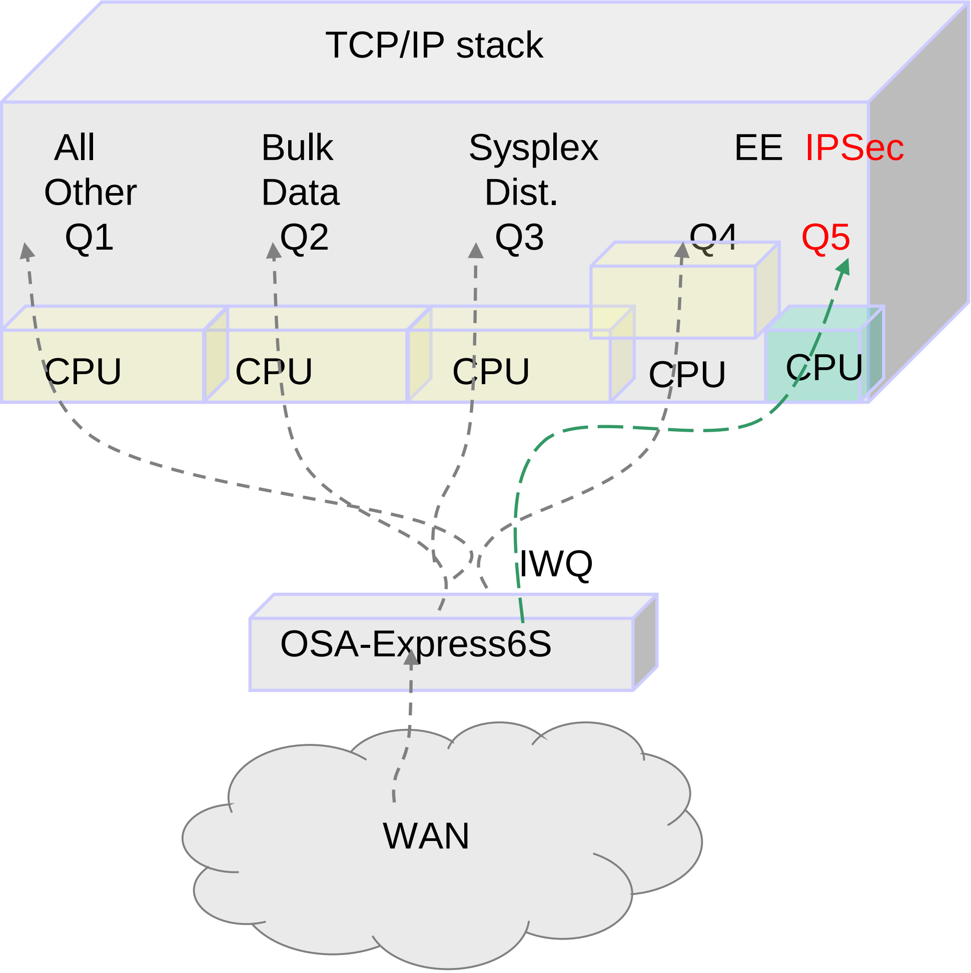
- ✓ IWQ IPSEC Queue Performance
- ✓ zERT Enablement
- ✓ Crypto Express Enhancements
- ✓ AT-TLS (Short & Long Handshake) Performance
- ✓ IP Security Using IPSec

V2R3 NETWORK SECURITY PERFORMANCE

2.1 IWQ IPSec Queue Performance

2.1a QDIO IWQ for IPSec (OSA Exp6S)

- New ancillary input queue for IPSec
- IPSec traffic serviced on its own processor (implies its own worker threads or CPU)
- Processing of IPSec queue is optimized since the only traffic on the queue is IPSec
- IPSec-protected (all traffic) such as bulk, SD, or EE traffic uses IPSec queue
- IWQ IPSec enables with IWQ using INTERFACE configuration option 'INBPERF DYNAMIC WORKLOADQ'
- IWQ IPSec is only available on the OSA Exp6s



V2R3 NETWORK SECURITY PERFORMANCE…

2.1 IWQ IPSec Queue Performance

2.1b Workload definitions for performance measurements

- RR10 1K/1K – 10 TCP connections sending and receiving 1000 bytes (persistent connections)
- STR3 20M/1 – 3 TCP streaming connections sending 20,000,000 bytes and receiving 1 byte
- CRR20(64/8k) – 20 TCP connect request response connections sending 64 bytes and receiving 8192 bytes
- All results collected on z14 or z13 running z/OS V2R3
- Results obtained using very lightweight applications – no application logic
- 2 CPs per LPAR
- IPSec encryption algorithm = AES_GCM_16 Keylength 128
- Dedicated OSAs (OSA Exp6S 10Gb)
- Tests with IWQ IPSec (new queue) enabled on Server side only

V2R3 NETWORK SECURITY PERFORMANCE…

2.1 IWQ IPSec Queue Performance

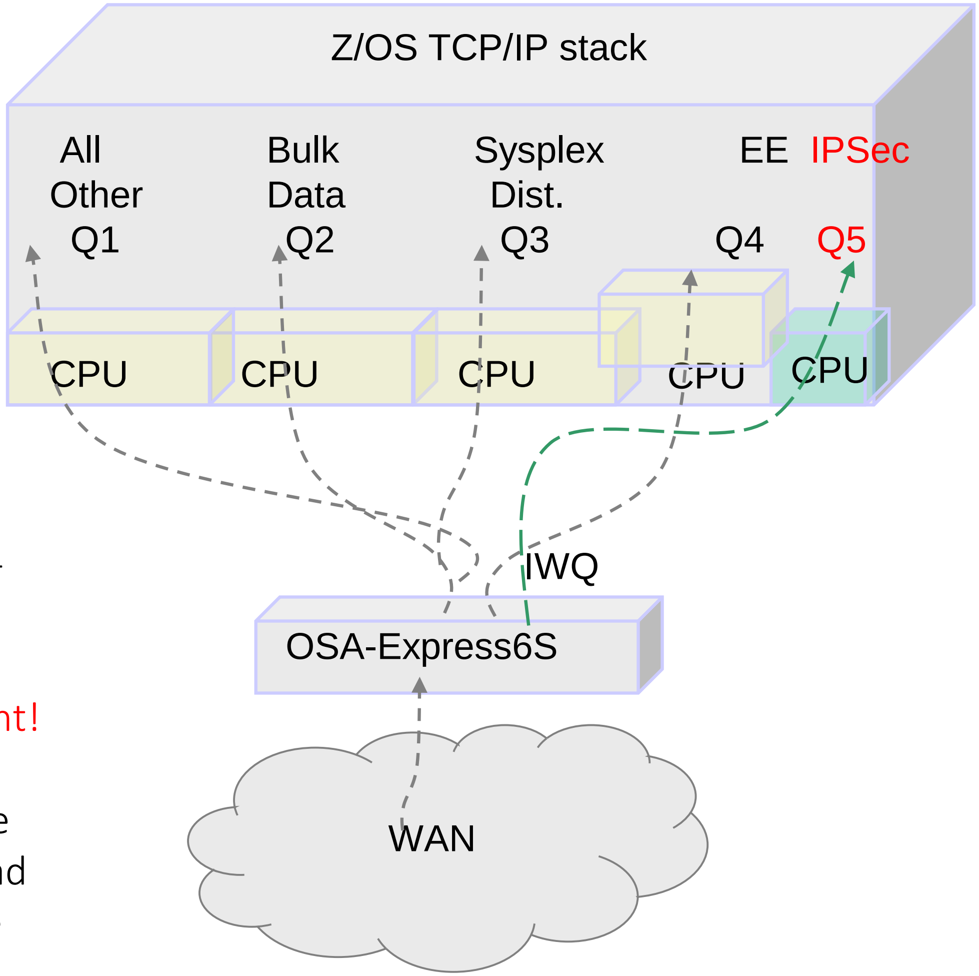
2.1c Performance Summary

Performance testing shows the following throughput improvement:

RR10 1K/1K encrypted text mixed with RR10 1K/1K clear text

	encrypted	/	clear
Without IWQ:	59 MB/Sec	/	53 MB/Sec
With IWQ:	74 MB/Sec	/	81 MB/Sec
	25 %	/	52 % improvement!

Moving encrypted workload to its own input queue improves the throughput of the encrypted workload by 25% while also increasing the throughput of the clear text workload by 52%

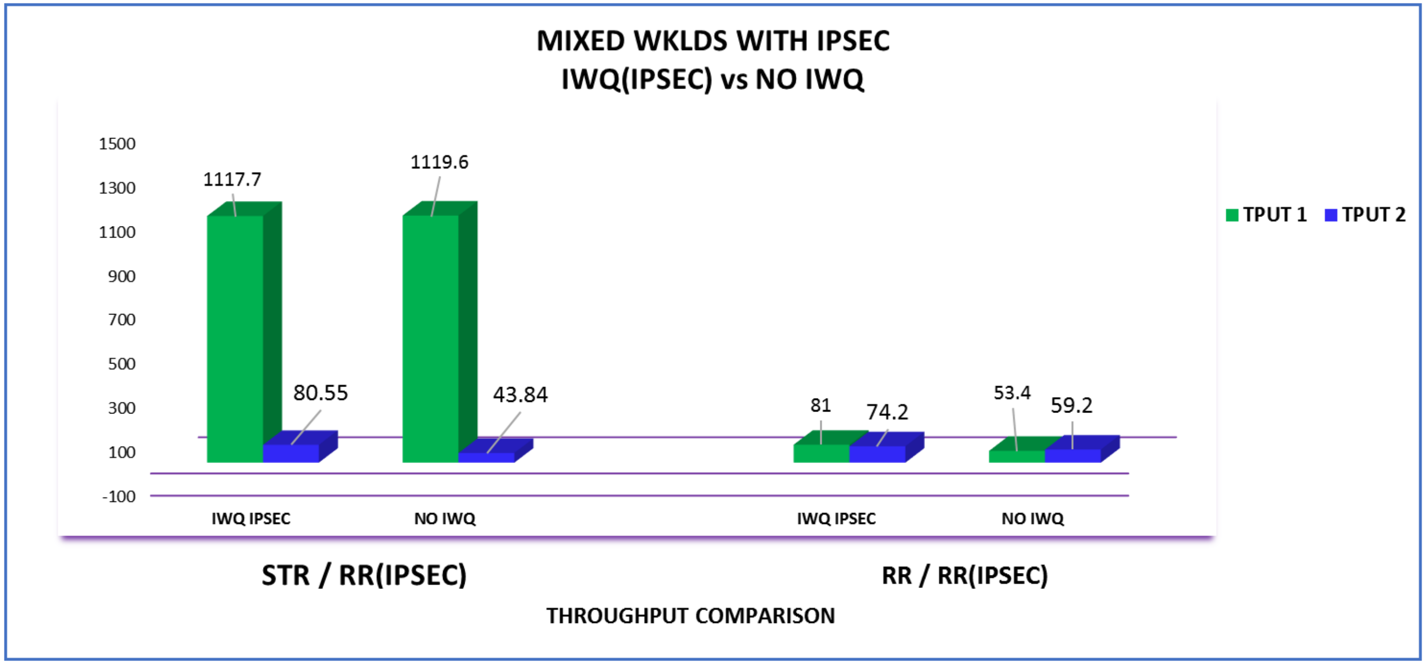


V2R3 NETWORK SECURITY PERFORMANCE…

2.1 IWQ IPSec Queue Performance

2.1d IWQ IPSec Queue Results (Mixed workloads)

- STR / RR(IPSec):
  - OSA Exp6s IPSec IWQ vs no IWQ for Mixed workloads (RR and STR)
  - RR traffic encrypted, STR in the clear
  - IPSec RR shows 80% throughput improvement mixed with STR (80 vs 44)
- RR / RR(IPSec):
  - IPSec RR and clear text RR both improve with IWQ IPSec (right side)



Note: Throughput for streaming and request response workloads are represented in MB/Sec

V2R3 NETWORK SECURITY PERFORMANCE…

2.1 IWQ IPSec Queue Performance

2.1e Useful Information

- OSA Requirement
  - OSA-Express6S Ethernet feature in QDIO mode running on an IBM z14 or later server
- Each Ancillary queue consumes
  - Approximately nine additional 4K pages of ECSA (36K DLC structures)
  - An additional but tunable amount of fixed 64-bit CSM as specified by the READSTORAGE parameter on the Interface statement in the TCP/IP configuration
- CSM Storage for Input Queues

  4k Input Buffers (4MB per input queue, fixed CSM HVCommon)

  Work Elements (fixed CSM ECSA)

CSA storage for control structures (static queues, CSM headers, etc.)

Estimate ("typical" 64bit fixed storage requirements) per OSA INTERFACE:

Without IWQ: 7MB (4MB + 2MB cache CSM HVC + ~ 1MB CSM HVC SPACs))

With IWQ: 24MB[1] (20MB + 2MB cache CSM HVC + ~ 2MB CSM HVC (SPACs))

Note: Here HVC is used as an abbreviation for HVCommon

With IWQ an additional ~ 17MB of CSM per OSA interface (most is HVCOMMON)

(e.g. for 4 OSA interfaces with IWQ ~ 68MB additional storage)

V2R3 NETWORK SECURITY PERFORMANCE…

2.1 IWQ IPSec Queue Performance

2.1f Enablement (PTF Install)

- For customers who already use IWQ with OSA-Express6S and apply the IWQ IPSec enablement PTFs, the IWQ IPSec function (input queue) will automatically be enabled (input queue is defined).
- The enablement will define the new input queue to OSA, but the new input queue will not be used (backed by 4MB of storage and TCP connection registered with OSA) until the first IPSec tunnel is activated.
- There are no configuration options for controlling each input queue type. The bulk queue is always active and the remaining IWQ input queues are used when the corresponding function is enabled (SD, EE and IPSec).
- Note – IWQ IPSec available on z/OS CS V2R2 and V2R3 via APARs P177649 and OA52275 and it requires OSA Express6S

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2a zERT Overview

z/OS® Encryption Readiness Technology (*zERT*) is a new capability provided by the *z/OS V2R3* Communications Server.

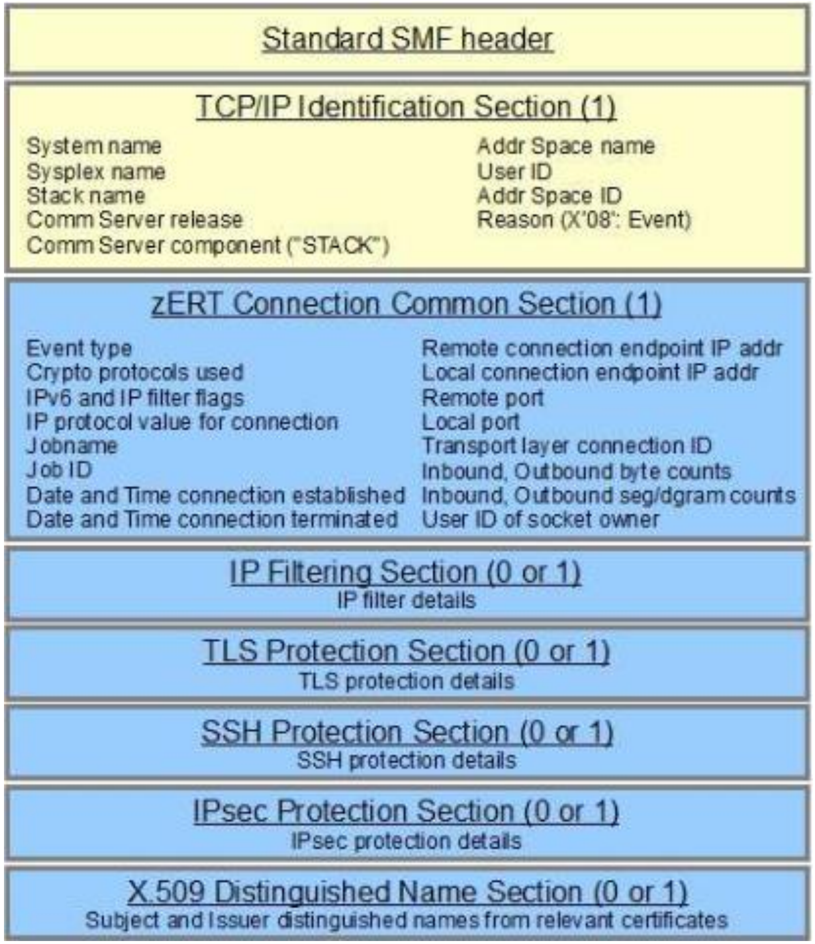
- zERT positions the TCP/IP stack as a central collection point and repository for cryptographic protection attributes for:
  - TCP connections that are protected by TLS/SSL and IPSEC or are unprotected
  - Enterprise Extender connections that are protected by IPSEC or are unprotected
- Reported through new SMF 119 records:
  - SMF 119 subtype 11 records from zERT Discovery function
  - SMF 119 subtype 12 records from zERT Aggregation function
  - Via SMF and/or new real-time Communications Server NMI services
- Function can be dynamically enabled and disabled
- IBM Security zSecure Audit V2.3 support zERT SMF subtype 11 records
  - Reporting
  - Forwarding to SIEM (Security Information and Event Management) like QRadar

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2b zERT Discovery Function

- Available in z/OS CS V2R3
- Attributes are collected and recorded at the connection level
- Generates SMF 119 subtype 11 “zERT Connection Detail” event records
- These records describe the cryptographic protection history of each TCP and EE connection
- Can generate large number of records depending on your z/OS system’s traffic patterns



Standard SMF header

TCP/IP Identification Section (1)

System name
Sysplex name
Stack name
Comm Server release
Comm Server component ("STACK")

Addr Space name
User ID
Addr Space ID
Reason (X'08': Event)

zERT Connection Common Section (1)

Event type
Crypto protocols used
IPv6 and IP filter flags
IP protocol value for connection
Jobname
Job ID
Date and Time connection established
Date and Time connection terminated

Remote connection endpoint IP addr
Local connection endpoint IP addr
Remote port
Local port
Transport layer connection ID
Inbound, Outbound byte counts
Inbound, Outbound seg/dgram counts
User ID of socket owner

IP Filtering Section (0 or 1)
IP filter details

TLS Protection Section (0 or 1)
TLS protection details

SSH Protection Section (0 or 1)
SSH protection details

IPsec Protection Section (0 or 1)
IPsec protection details

X.509 Distinguished Name Section (0 or 1)
Subject and Issuer distinguished names from relevant certificates

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2c zERT Aggregation Function

- Available in z/OS CS V2R3
- Connection level attributes are aggregated by security session
  - Server IP address and port
  - Client IP address
- Generates SMF 119 subtype 12 “zERT Summary” records on regular intervals
- These records describe the repeated use of security sessions over time
- Aggregation can greatly reduce the volume of SMF records while maintaining the fidelity of the information
- Well suited for reporting applications

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2d zERT Performance Consideration

- Ideally, customers will choose to run with recording of subtype 11 records disabled and only record subtype 12 records to SMF for consumption by zERT visualization.
- However, the existence of zSecure support for subtype 11 records may cause more customers to enable subtype 11 recording.
  - Performance testing so far has revealed the bulk of performance impact from zERT is in the discovery portion and not in the subtype 11 creation.
  - Expectations of the aggregation function should be in record reduction (elimination of subtype 11) and not in zERT performance impact.

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2e z/OS mechanisms to protect TCP/IP traffic

z/OS provides 4 mechanisms to protect TCP/IP traffic:

1 TLS/SSL direct usage

- Application is explicitly coded to use these
- Configuration and auditing is unique to each application
- Per-session protection
- TCP only

2 Application Transparent TLS (AT-TLS)

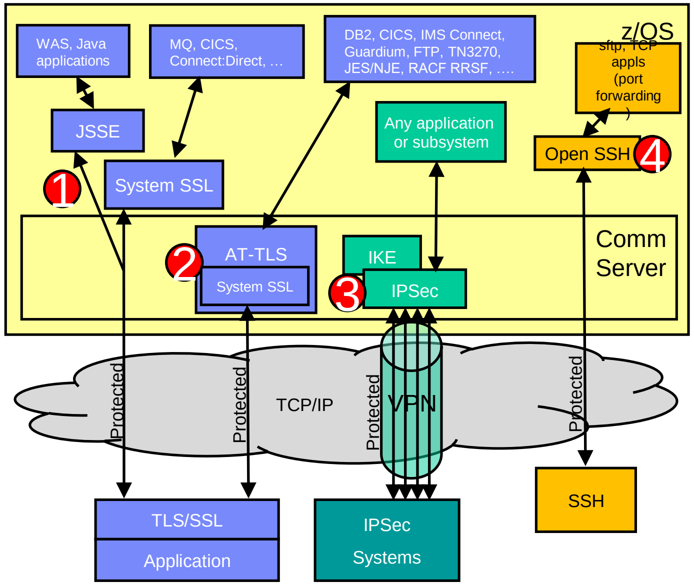
- TLS/SSL applied in TCP layer as defined by policy
- Configured in AT-TLS policy via Configuration Assistant
- Auditing through SMF 119 records
- Typically transparent to application
- TCP/IP stack is user of System SSL services

3 Virtual Private Networks using IPSec and IKE

- “Platform to platform” encryption
- IPSec implemented in IP layer as defined by policy
- Auditing via SMF 119 records at tunnel level only
- Completely transparent to application
- Wide variety (any to all) of traffic is protected
- IKE negotiates IPSec tunnels dynamically

4 Secure Shell using z/OS OpenSSH

- Mainly used for sftp on z/OS, but also offers secure terminal access and TCP port forwarding
- Configured in ssh configuration file and on command line
- Auditing via SMF 119 records
- TCP only



V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2f zERT Discovery and Aggregation

- zERT Discovery **– available at V2R3 General Availability**
  - Attributes are collected and recorded at the connection level
  - SMF 119 subtype 11 “zERT Connection Detail” records
  - These records **describe the cryptographic protection history of each TCP and EE connection**
  - Measures are in place to minimize the number of subtype 11 records, but they could still be very voluminous – Also, see next bullet
- zERT Aggregation **– Available in V2R3 with Apar PI83362 / UI54759**
  - Attributes collected by zERT discovery are aggregated by security session
  - SMF 119 subtype 12 “zERT Summary” records
  - These records **describe the repeated use of security sessions over time**
    - **Aggregate connection data from repeated connections between a TCP client and server**
  - Aggregation can greatly reduce the volume of SMF records while maintaining the fidelity of the information – well suited for reporting applications

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Deployment on z/OS

2.2g Test Environment

- Hardware
  - z13 model 2964-760 with four LPARs, each with
    - 64G central storage
    - up to four dedicated general purpose CPs and up to two zIIP processors
    - One OSA Express 5S 10 Gbe adapter
    - One Crypto Express5S adapter (co-processor mode)
  - CPs and adapters are configured as dedicated

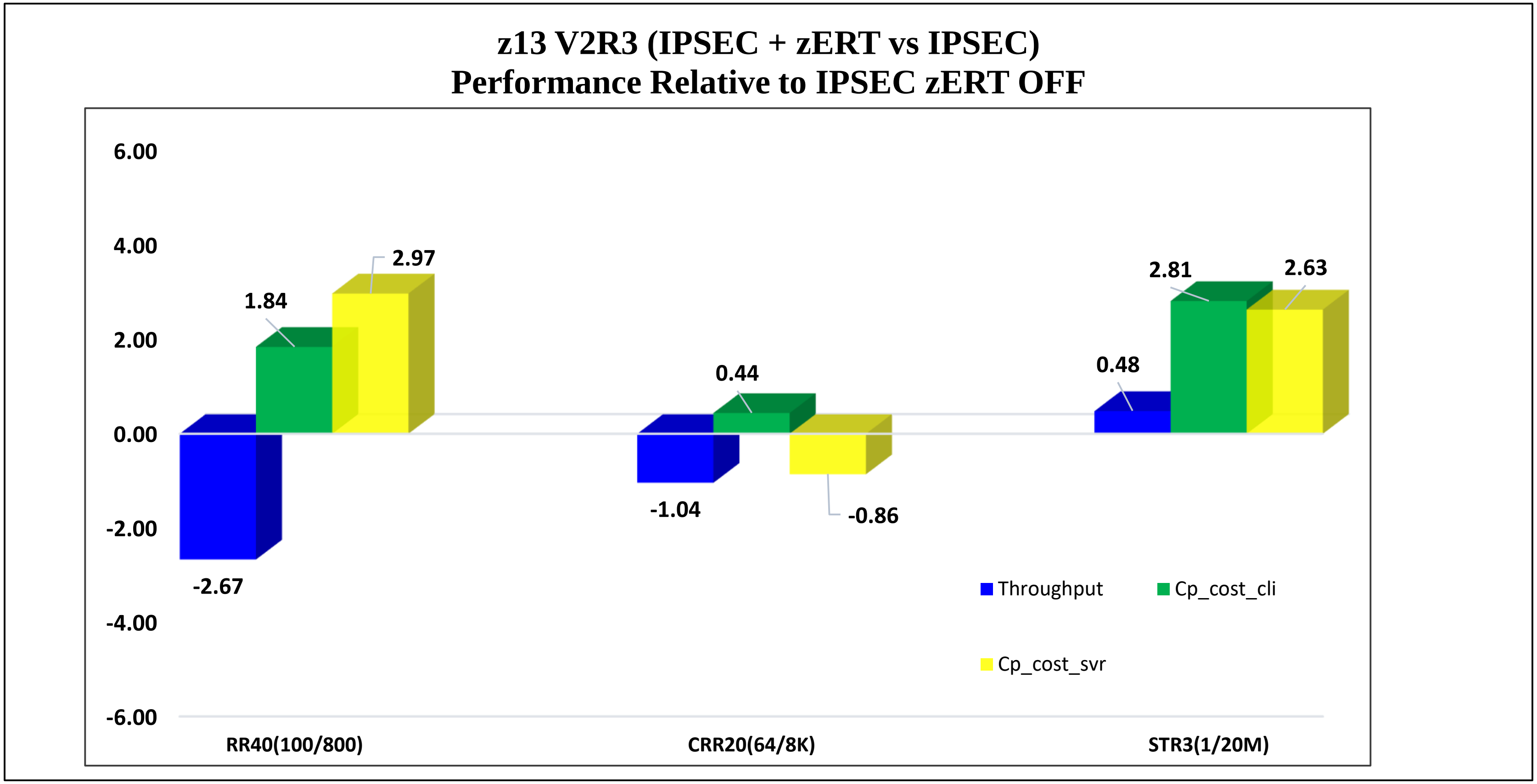
- Software
  - z/OS V2R3
  - ICSF FMID: HCR77C1
  - IPSec (AES_GCM_16 KeyLength 128, ESP NULL)
  - Pre Shared key
  - Network configuration: 10GbE, Jumbo frames, Segmentation offload enabled, IWQ enabled

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2h Enabling zERT for IPSec Workloads

Enabling zERT for IPSEC workloads – No SMF records



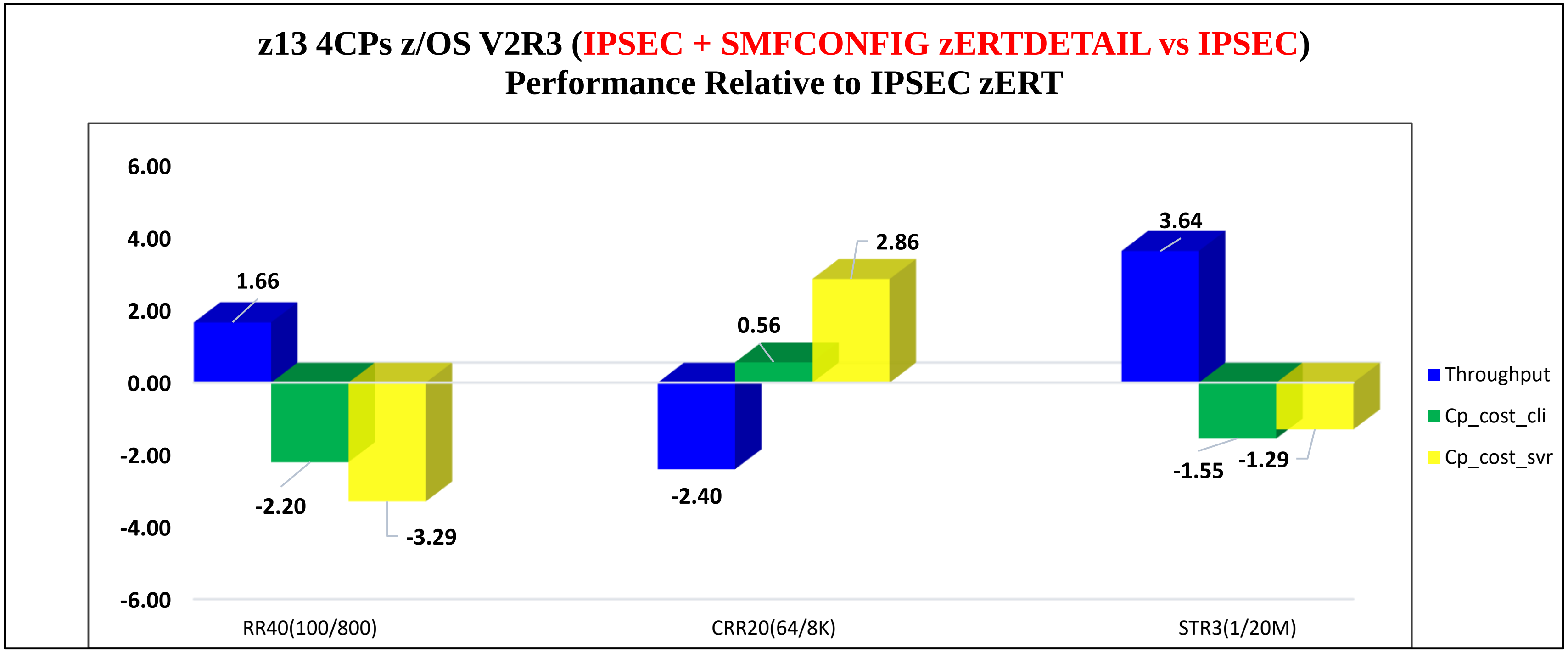
Note: Chart shows the effect of enabling zERT for IPSec workloads on throughput and CPU cost as percent increase/decrease

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2h Enabling zERT for IPSec Workloads …

Enabling zERT for IPSEC workloads – SMF subtype 11 records



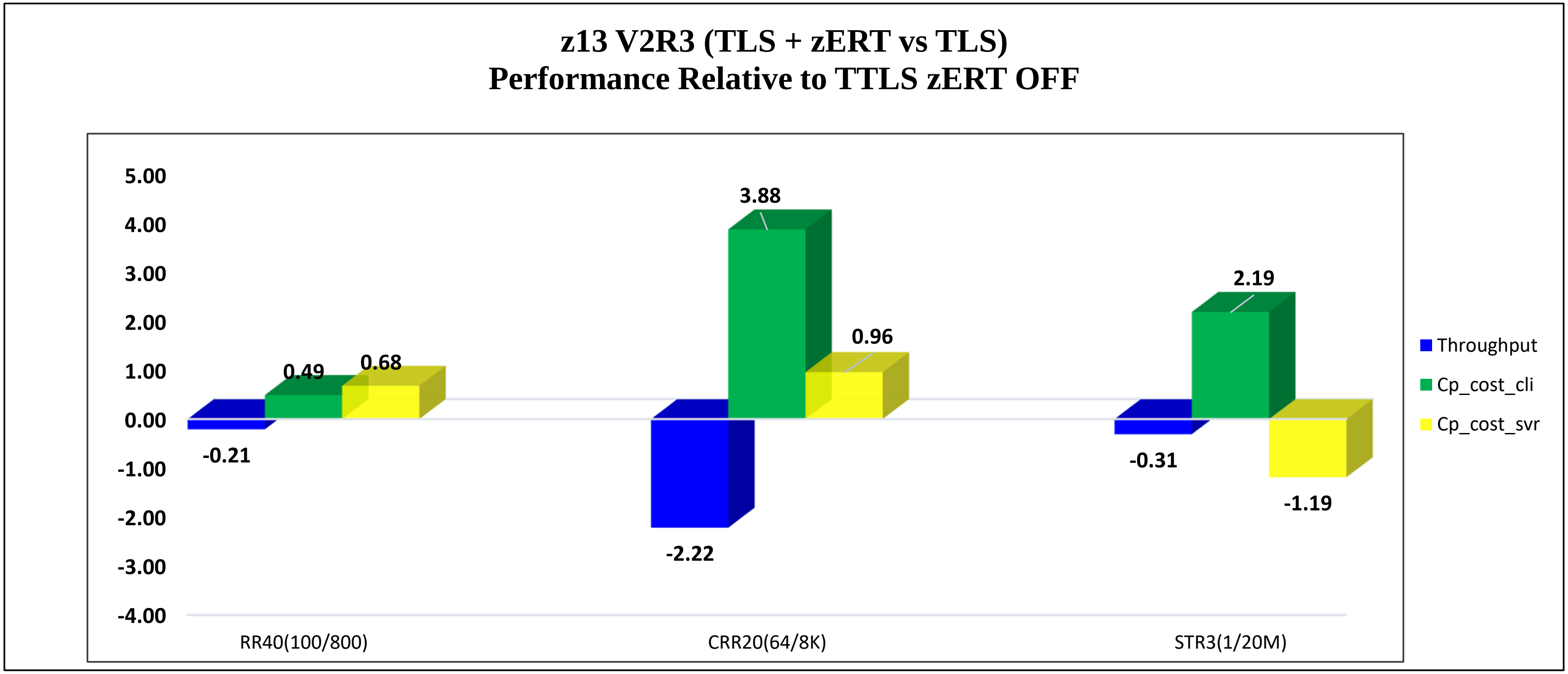
Note: Chart shows the effect of enabling zERT (zERTDETAIL) for IPSec workloads on throughput and CPU cost as percent increase/decrease

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2i zERT Enablement for TLS Workloads

Enabling zERT for TLS Workloads



Note – All TLS measurements were done using AT-TLS

V2R3 NETWORK SECURITY PERFORMANCE…

2.2 zERT Enablement Performance

2.2j zERT Discovery and Aggregation - Overall summary

- zERT Discovery & Aggregation
  - Enabling zERT has little to no impact on latency or CPU consumption
  - The CPU results reflect networking related CPU costs only which are a small fraction of the overall system CPU costs
  - Results obtained using applications with no application logic (micro-benchmarks)
  - In a real workload the percent of CPU increases or decreases would be much smaller compared to the overall system CPU utilization
  - All zERT storage obtained from 64-Bit private (minimize footprint)
  - Aggregation will minimize SMF records created
    - Attributes collected by zERT discovery are aggregated by security session
    - These records describe the repeated use of security sessions over time
    - Aggregate connection data from repeated connections between a TCP client and server
    - Aggregation can greatly reduce the volume of SMF records while maintaining the fidelity of the information – well suited for reporting applications

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3a Overview

- How will enabling network security affect the performance of your workloads?
  - Impacts to latency and throughput
  - CPU consumption
- How do network traffic patterns affect network encryption performance and overhead?
- How can you optimize performance for network encryption?
- What performance enhancements does z14 offer for encryption?

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3b Benchmark Environment

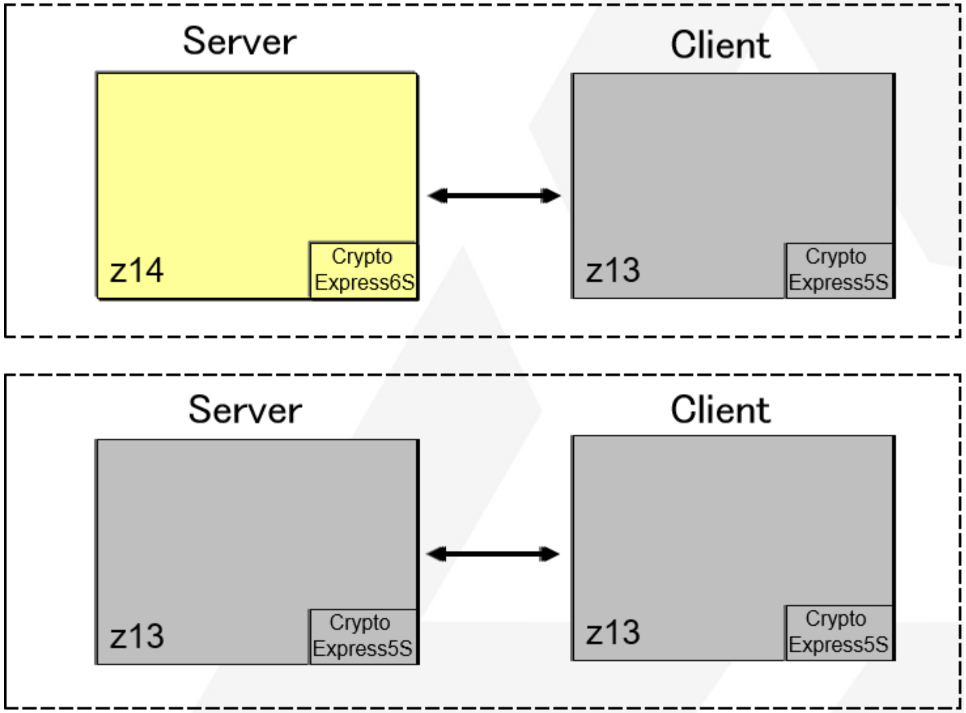
- Hardware
  - z14 model 3906-785 with four LPARs, each with
    - 64G central storage
    - up to four dedicated general purpose CPs and up to two zIIP processors
    - One each: OSA Express6S & 5S 10 Gbe adapter
    - One each: Crypto Express6S & 5S adapter (co-processor mode)
  - z13 model 2964-760 with four LPARs, each with
    - 64G central storage
    - up to four dedicated general purpose CPs and up to two zIIP processors
    - One OSA Express 5S 10 Gbe adapter
    - One Crypto Express5S adapter (co-processor mode)
  - CPs and adapters are configured as dedicated

- Software
  - z/OS V2R3
  - CICS V5R1 (CICS Sockets Environment OTE=YES)
  - ICSF FMID: HCR77C1
  - TLS V1.2 (TLS_RSA_WITH_AES_128_GCM_SHA256)
  - RSA keyring size was 2K
  - Network configuration: 10GbE, Jumbo frames, Segmentation offload enabled, IWQ enabled

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3c Test Environment



- All measurements performed with z/OS as both the client and the server:
- The focus of the benchmarks and the results shown reflect the **server side (z14)**
- The client side configuration and HW/SW was identical
- The server side configuration was identical but varied between z14 and z13

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

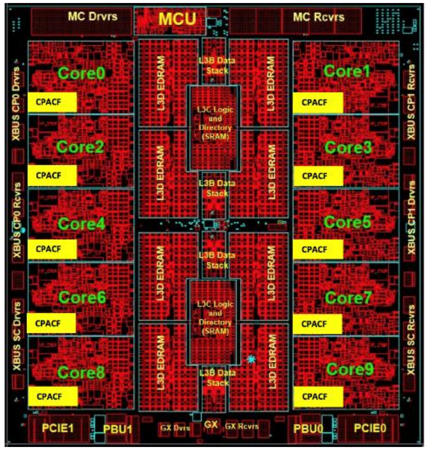
2.3d Integrated Cryptographic Hardware

CP Assist for Cryptographic Functions (CPACF)

- Hardware accelerated encryption on every microprocessor core
- Performance improvements of up to 7x for selective encryption modes

Crypto Express6S

- Next generation PCIe Hardware Security Module (HSM)
- Performance improvements up to 2x
- Industry leading FIPS 140-2 Level 4 Certification Design



Why is it valuable:

- More performance = lower latency + less CPU overhead for encryption operations
- Highest level of protection available for encryption keys
- Industry exclusive "protected key" encryption

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

TLS/SSL using AT-TLS

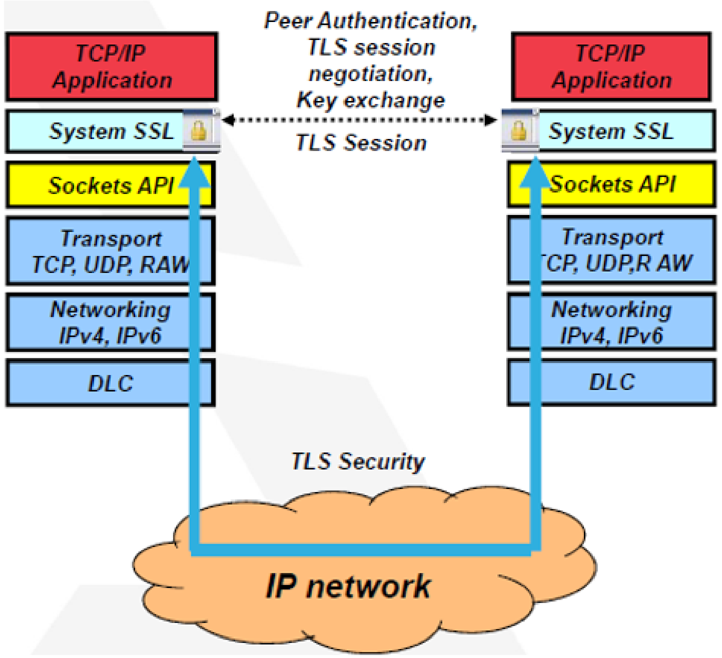
V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3e TLS/SSL Overview

- Transport Layer Security (TLS) is an IETF standard based on Netscape's old proprietary Secure Sockets Layer (SSL) protocol
  - Current version is TLSv1.3 (recently approved) but TLSv1.2 is predominantly used
- TLS traditionally provides security services as a socket layer service
  - Applications must be modified to call these services
- TLS requires a reliable transport protocol (TCP)
- z/OS supports two complete TLS/SSL implementations
  - z/OS Cryptographic Services System SSL
  - Java Secure Sockets Extension (JSSE)

- However, there is an easier way …
  - Application Transparent TLS (AT-TLS)

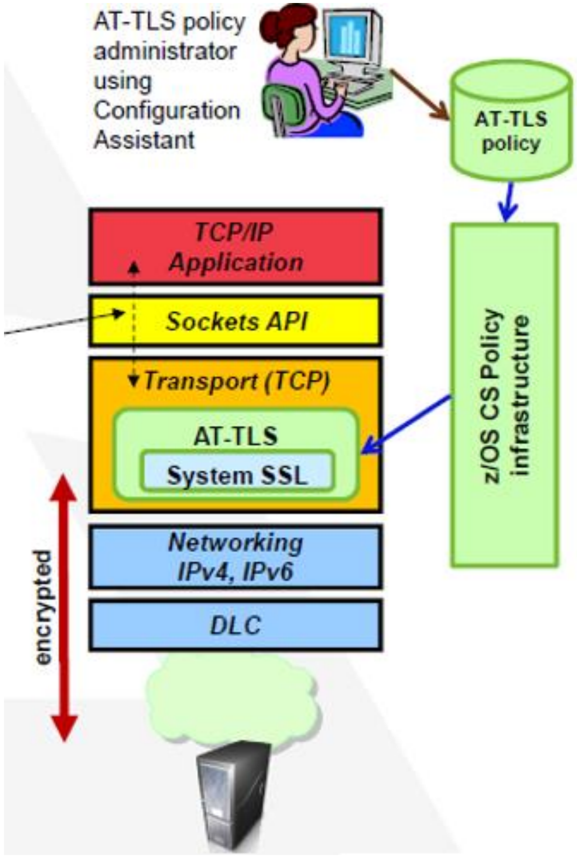


V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3f AT-TLS Overview

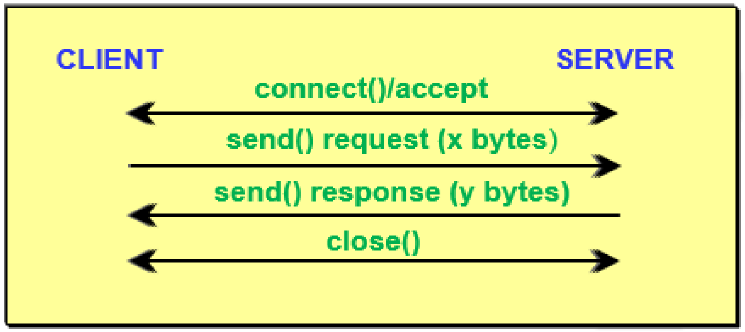
- Policy-based TLS in the TCP/IP stack
  - TLS process performed in TCP layer without any application change
  - AT-TLS policy specifies which TCP traffic is to be TLS protected based on selected criteria
    - local address/port, remote address/port, z/OS userid or jobname, …
- Application transparency
  - Can be fully transparent to application
  - An optional API allows applications to inspect/control aspects of AT-TLS processing ("application-aware" and "application-controlled")
- Available to TCP applications
  - Supports all programming languages except PASCAL
- Supports all standard configurations
  - z/OS as a client or server
  - Server authentication (server identifies self to client)
  - Client authentication (both sides identify selves to each other)
- Relies on System SSL for TLS processing
  - Remote endpoint sees RFC-compliant implementation



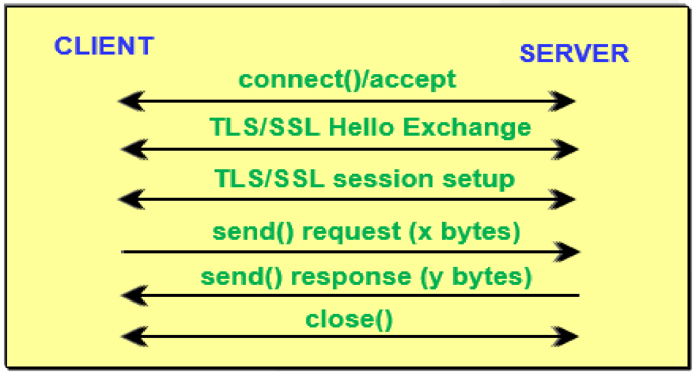
V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3g Non-Persistent Connections and TLS/SSL



With TLS/SSL – Full Handshakes



- CRRz(x/y) micro-benchmarks:
  - CRR describes a connect-request-response workload where
    - z is the number of client-server tasks
    - Each client connects to its server, sends x bytes, receives y bytes from the server, closes the connection, and process is repeated
    - Models non-persistent connection request/response traffic patterns such as web server traffic

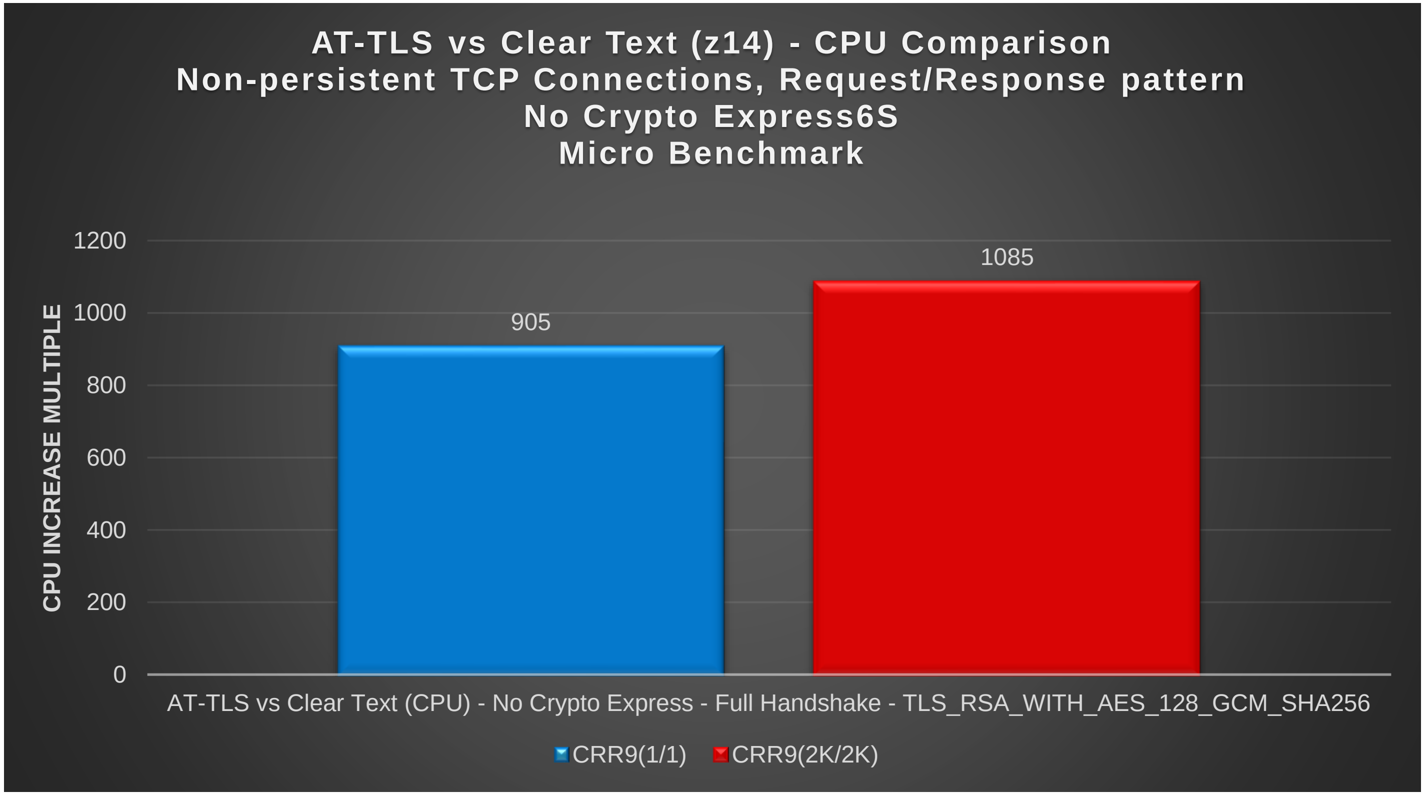
Notes:

- Micro-benchmark includes all z/OS host networking-related CPU overhead (up through the application socket layer)
- But the synthetic socket applications have **no** application logic
  - The networking related CPU cost equals entire application CPU cost
  - In real workloads, networking related CPU cost is fraction of overall application transaction cost (these benchmarks show the worst case scenario from a networking related CPU perspective)
- **2 full round-trip network flows** are needed before the client receives the reply (i.e. transaction latency)
- A full TLS/SSL handshake **doubles** the roundtrip network flows

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3h AT-TLS with no Crypto Express Coprocessor



TLS/SSL protection for non-persistent TCP connections with a single request/response pattern can introduce significant CPU overhead

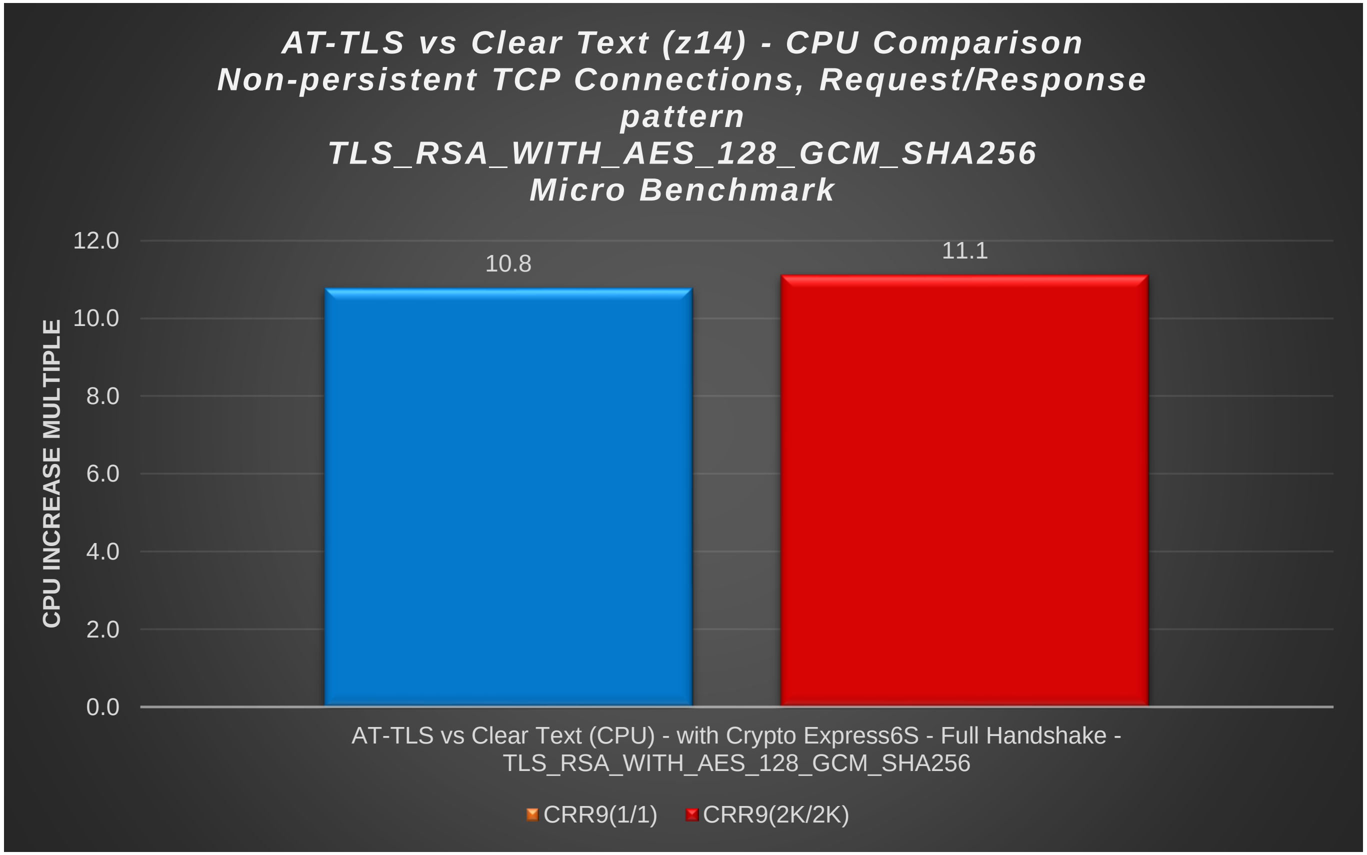
- *1085 times the CPU overhead of a clear text TCP connection in this measurement!*
- *But remember, no Crypto Express Coprocessor*

So don't panic, it gets better!

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3i AT-TLS with Crypto Express6S Coprocessor



Adding a Crypto Express6S coprocessor significantly reduces the CPU overhead of performing asymmetric encryption processing for TLS/SSL handshakes

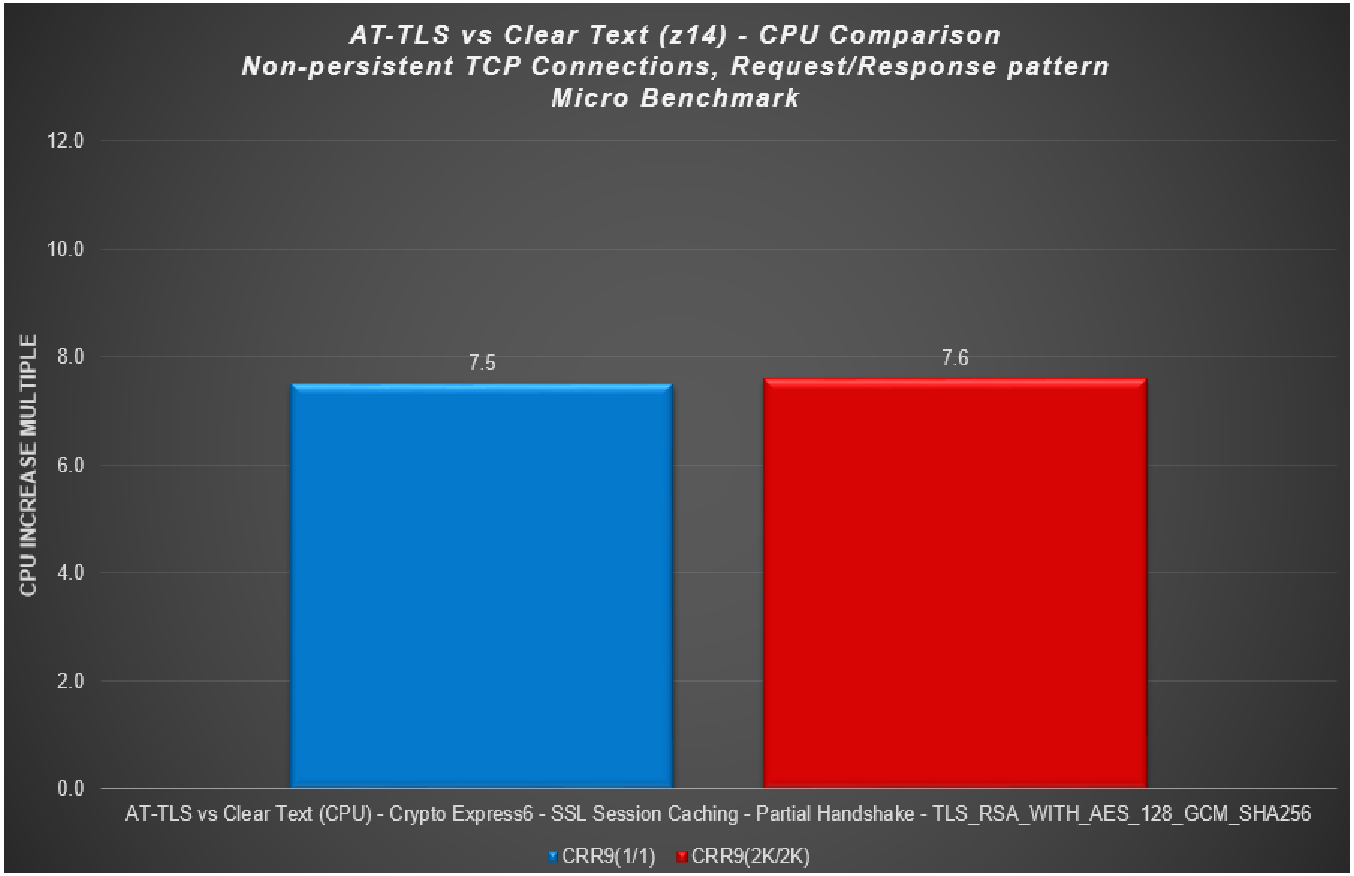
- *Reduces CPU consumption by over 99% (compared to benchmark with no Crypto Express6S coprocessor)*
- *Driving around 20,000 TCP connections and TLS handshakes per second*
- *CryptoExpress6S utilization (1 feature) around 75%*
- *But 11X CPU increase versus clear text is still significant**!***

Keep reading, it gets better…

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3j AT-TLS with Session Caching



Enabling AT-TLS and System SSL session caching allows repeated connections from same clients to perform an optimized TLS/SSL handshake

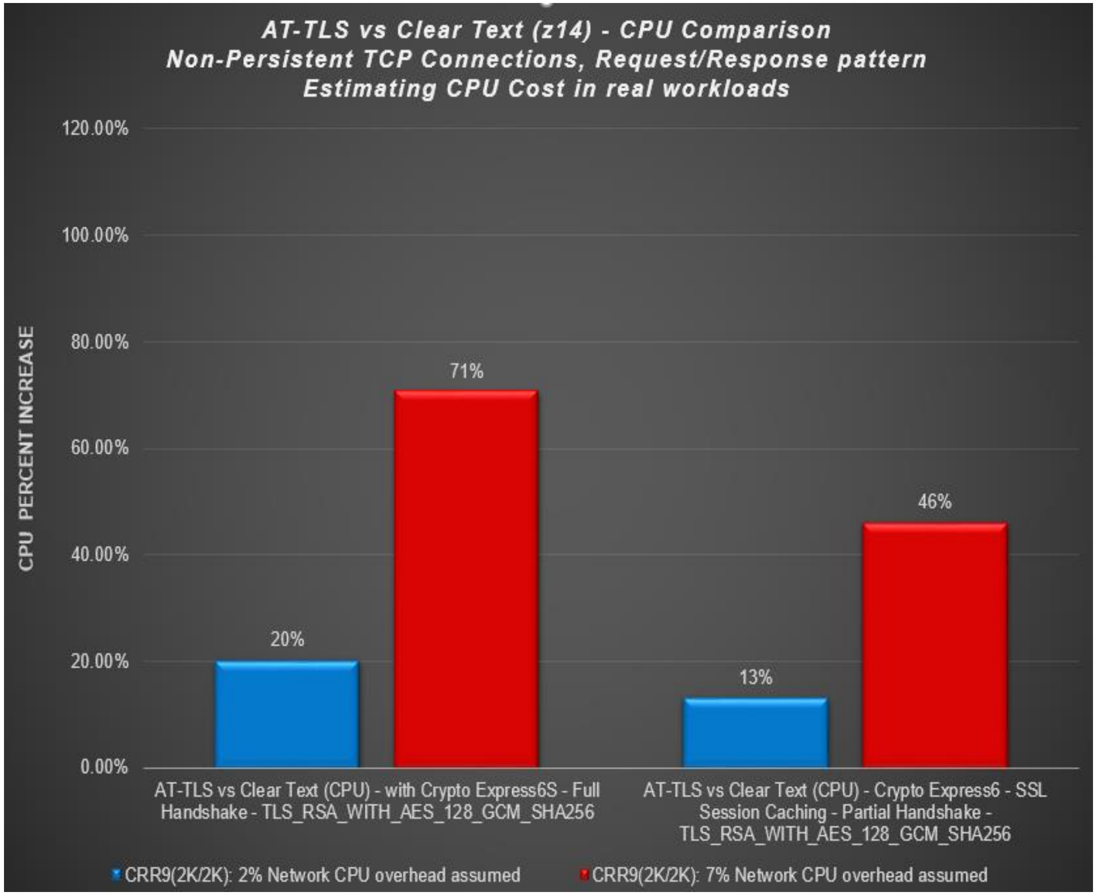
- *Reduces number of round-trip network flows needed for TLS/SSL handshakes to 1 (from 2 round-trips required for a full handshake)*
- *Avoids all the expensive asymmetric encryption processing needed to generate new session keys*
- *Avoids use of Crypto Express coprocessor (saves capacity for new session TLS handshakes)*
- *But 7.5X CPU increase versus clear text is still significant!*

The story gets even better…

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3k What is the cost for real workloads?



The networking CPU overhead (without encryption) is small portion of overall CPU overhead of a workload driven by non-persistent connections

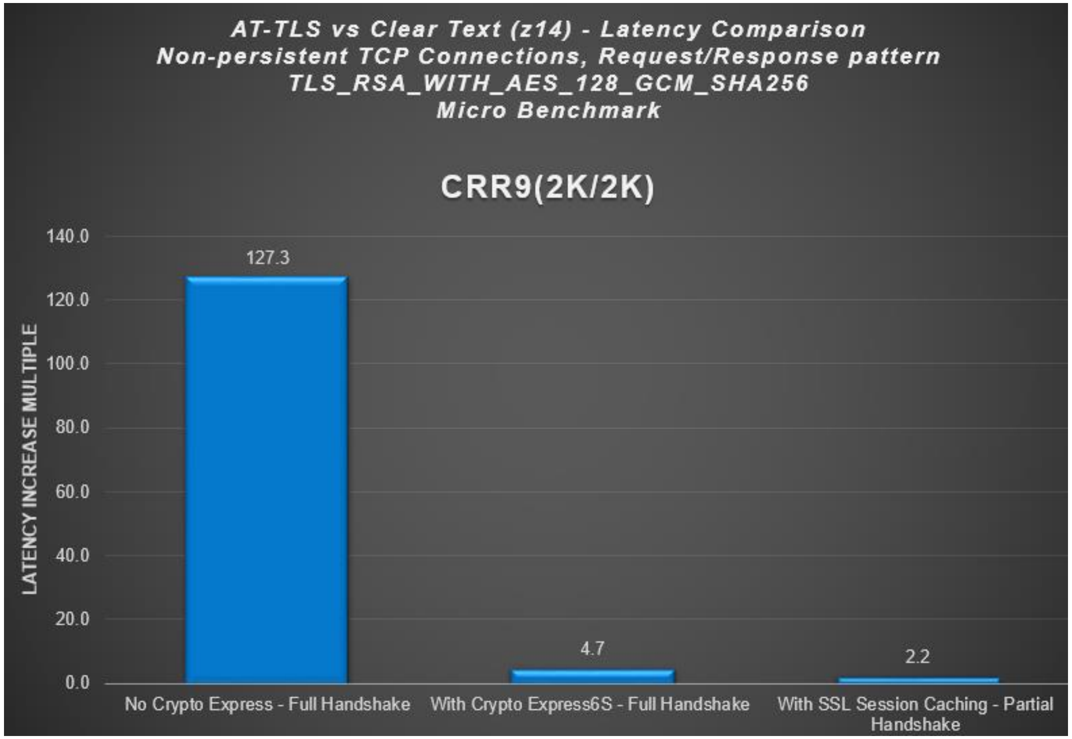
- Based on benchmark data for CICS synthetic workloads (i.e. no application logic), networking related CPU cost for processing non-persistent connections range from 5-10% of total cost
- Real applications with business logic will likely have lower percentage: we assumed 2-7% in this chart
  - Note: There are many variables that can affect CPU cost in user environments – these benchmarks were performed in controlled dedicated test environments and may not reflect performance attributes in your environment
- To perform a custom estimate for a specific workload, use the table below that shows the networking CPU cost from these benchmarks
  - CPU costs listed as milliseconds of CPU per connection
- These are z14 benchmarks – if comparing to older z processors, adjustments will be needed
  - CRR9 (2K/2K) benchmarks (AT-TLS with TLS_RSA_WITH_AES_128_GCM_SHA256)

z14 with Crypto Express6S • Full Handshake	z14 with Crypto Express6S • SSL Session Caching (100% cache hit)
0.294	0.209

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3l What is the impact on latency?



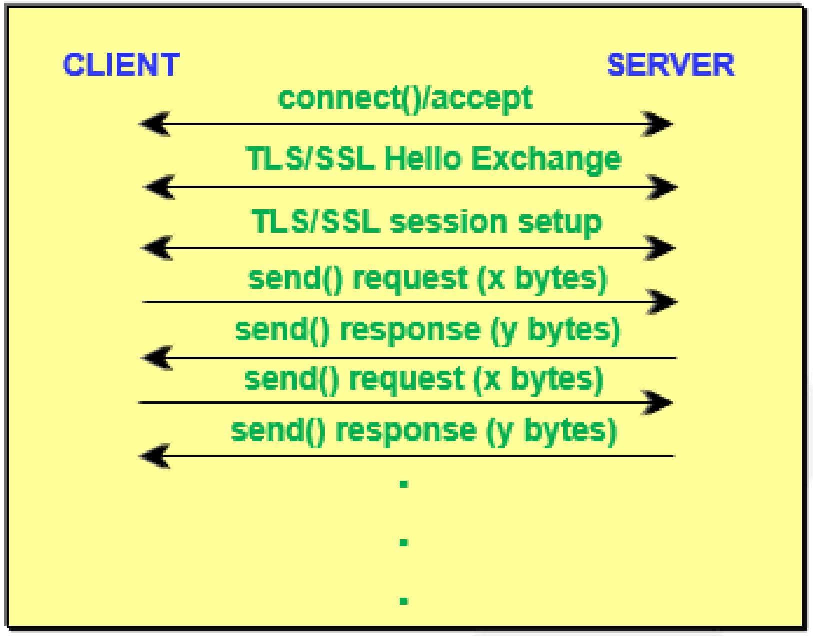
TLS/SSL processing for non-persistent connections introduces additional overhead for network encryption but also introduces additional network flows

- A full TLS/SSL handshake introduces 2 additional full round-trip network flows between client/server
- Worst case, doubling number of round-trip network flows can impact transaction latency
- Increase in latency will be significantly larger across high latency networks – these benchmarks had minimal network latency (same LAN)
- **Observations:**
  - Performing TLS/SSL handshakes for CRR workloads without Crypto Express should be avoided for latency sensitive or high workload volume applications
  - Crypto Express6S coprocessors significantly improve performance and latency – but note that 2 additional round-trip network flows can have a significant impact on latency
  - SSL/TLS Session Caching can dramatically improve performance and latency when most sessions found in the cache (however, 1 extra round-trip network flow still needed)

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3m Persistent Connections and TLS/SSL



RRz(x/y) micro-benchmarks:

- RR describes a connect-request-response workload where
  - z is the number of client-server tasks
  - Each client connects to its server, sends x bytes, receives y bytes from the server, and process is repeated (connection is not terminated after each request/response)
  - Models persistent connection request/response traffic patterns such as TN3270, CICS web services, DB2, etc.

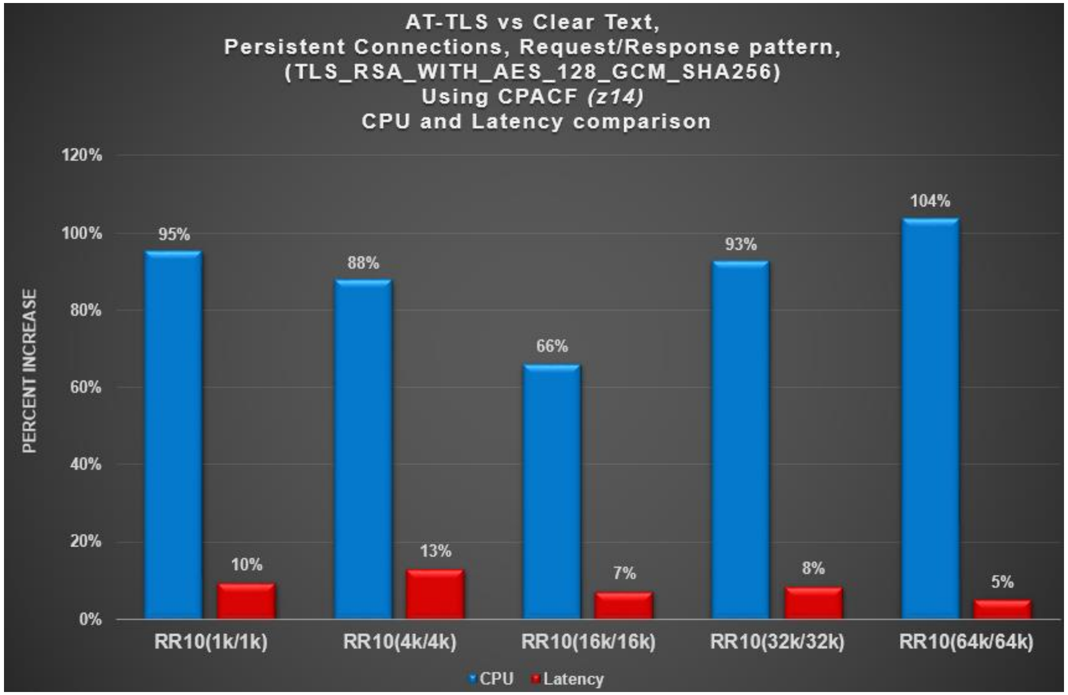
Notes:

- Micro-benchmark includes all z/OS host networking-related CPU overhead (up through the application socket layer)
- But the synthetic socket applications have *no* application logic
  - The networking related CPU cost equals entire application CPU cost
  - In real workloads, networking related CPU cost is fraction of overall application transaction cost (these benchmarks show the worst case scenario from a networking related CPU perspective)
- The TLS/SSL handshake occurs only at beginning of connections and is amortized across life of connection (much more efficient than non-persistent connections)
- Main cost of TLS/SSL becomes encrypt/decrypt operations on data flowing over connection

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3n Persistent Connections – CPU and Latency Impact



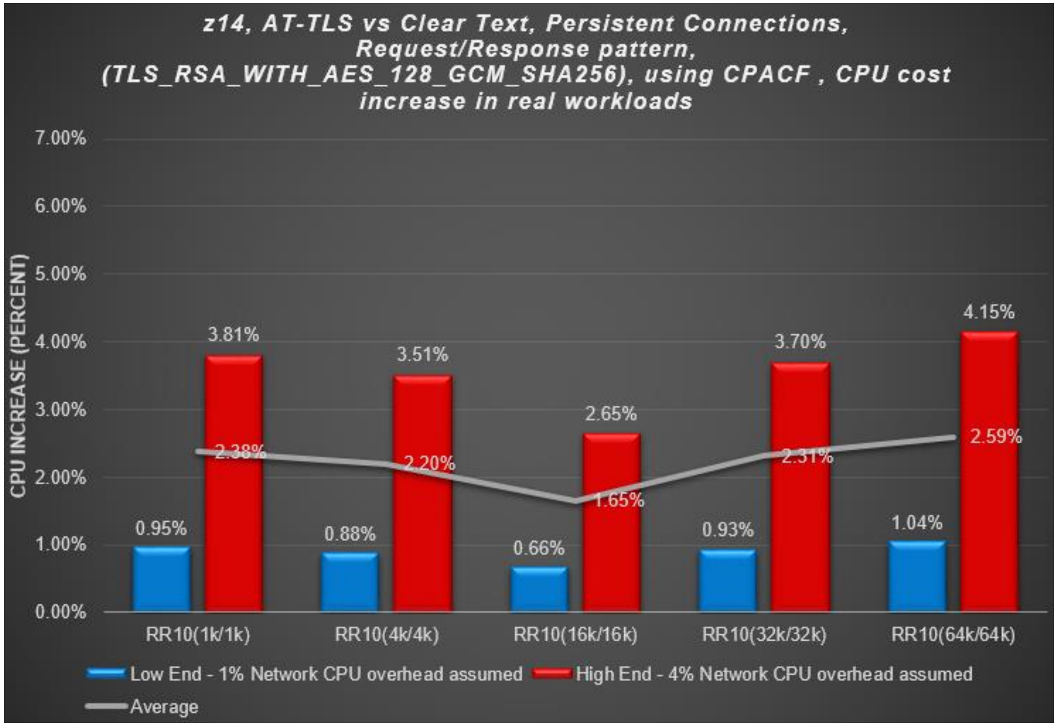
TLS/SSL processing for persistent TCP connections is much more efficient than non-persistent TCP connections as impact of TLS/SSL handshakes is largely eliminated

- TLS/SSL processing for persistent connections mainly consists of symmetric encryption/decryption operations that occur on z processors
- Cipher suites implemented on CPACF significantly improve performance – and z14 offers significant performance gains for selected cipher suites
- Observations:
  - Latency impact is fairly minimal as number of network flows is not changed in request/response traffic patterns
    - CPU impact is dependent on the size of the data (size of request and response)
  - Significant improvements for AES GCM performance on z14 CPACF – even for large data
  - Explore options to change workloads to use persistent TCP connections (significant CPU and latency benefits)

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3o Persistent Connections – Cost for real workloads



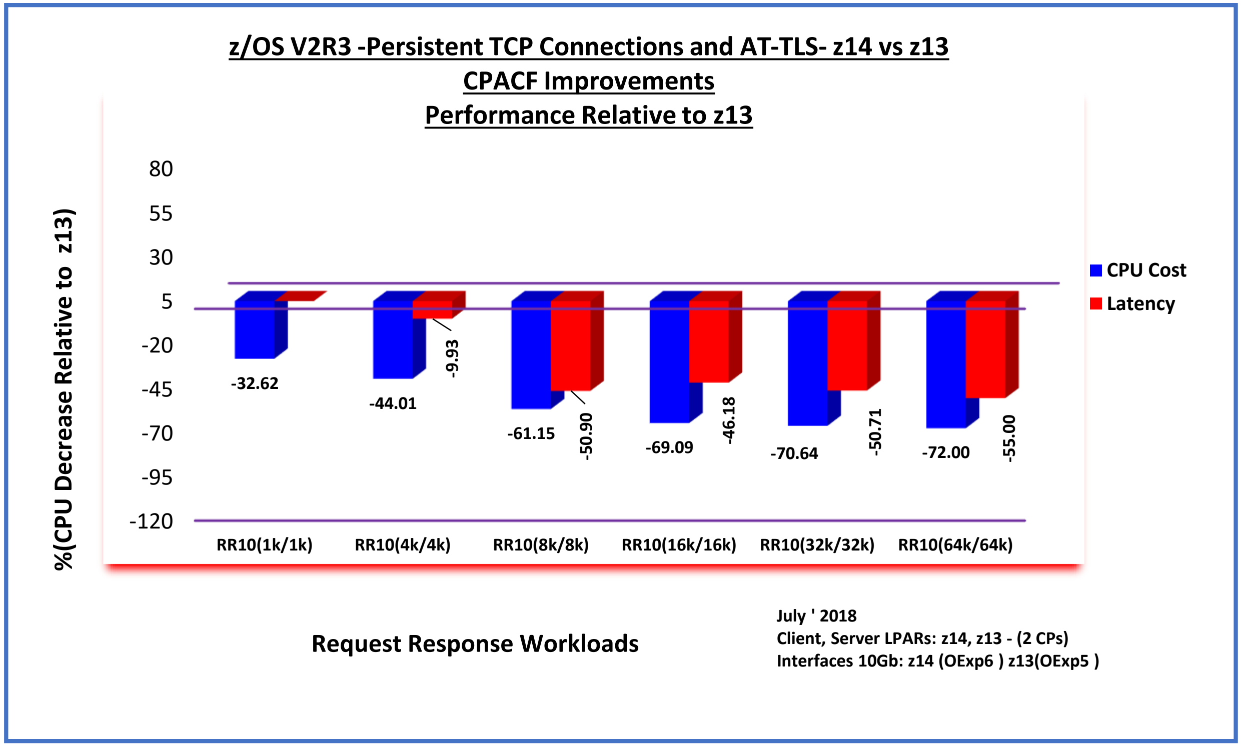
Data sizes	CPU cost
RR10(1k/1k)	0.014
RR10(4k/4k)	0.017
RR10(16k/16k)	0.026
RR10(32k/32k)	0.049
RR10(64k/64k)	0.091

The networking CPU overhead (without encryption) is small portion of overall CPU overhead of a workload driven by non-persistent connections

- Based on benchmark data for CICS synthetic workloads (i.e. no application logic), networking related CPU cost for processing persistent connections of about 2% of total cost
- Real applications with business logic will likely have lower percentage: we assumed 1-4% in this chart
  - Note: There are many variables that can affect CPU cost in user environments – these benchmarks were performed in controlled dedicated test environments and may not reflect performance attributes in your environment
- To perform a custom estimate for a specific workload, use the table below that shows the networking CPU cost from these benchmarks
  - CPU costs listed as milliseconds of CPU per request/response
- These are z14 benchmarks – if comparing to older z processors, adjustments will be needed
  - RR10 (various data sizes) benchmarks (AT-TLS with TLS_RSA_WITH_AES_128_GCM_SHA256)

V2R3 NETWORK SECURITY PERFORMANCE…

2.3p Persistent TCP Connections and AT-TLS - z14 vs z13 (CPACF improvements)



July ' 2018
Client, Server LPARs: z14, z13 - (2 CPs)
Interfaces 10Gb: z14 (OExp6) z13(OExp5)

Significant performance improvements for selected cipher suites in z14 CPACF

- Significant reduction in CPU cost – up to 72% lower CPU consumption per transaction
- Significant reduction in network latency for most data sizes (up to 55%)

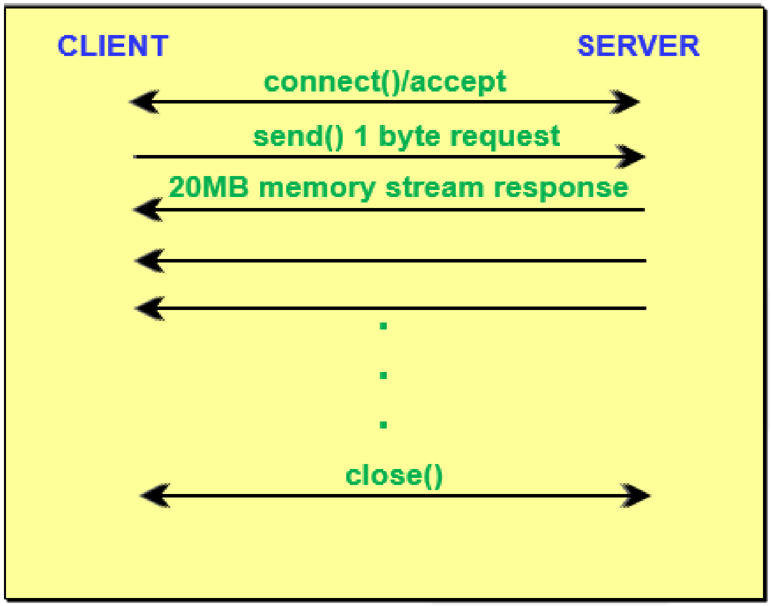
Recommendations:

- Evaluate performance benefits (CPU and latency) that z14 can offer your workloads compared to your current processors
  - Note: Performance benefits will be even more significant if you are migrating from older z processors
- Evaluate use of cipher suites that offer very compelling performance advantages on z14 CPACF (like AES GCM suites)

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3q Streaming Connections (TLS/SSL)



STRz(1/20M) micro-benchmarks:

- STR describes a streaming workload where
  - z is the number of client-server tasks
  - Each client connects to its server, sends 1 byte, receives 20 MB from the server, and the process is repeated (connection is not terminated after each 1 byte send, 20MB).
  - Models bulk data transfer traffic such as FTP

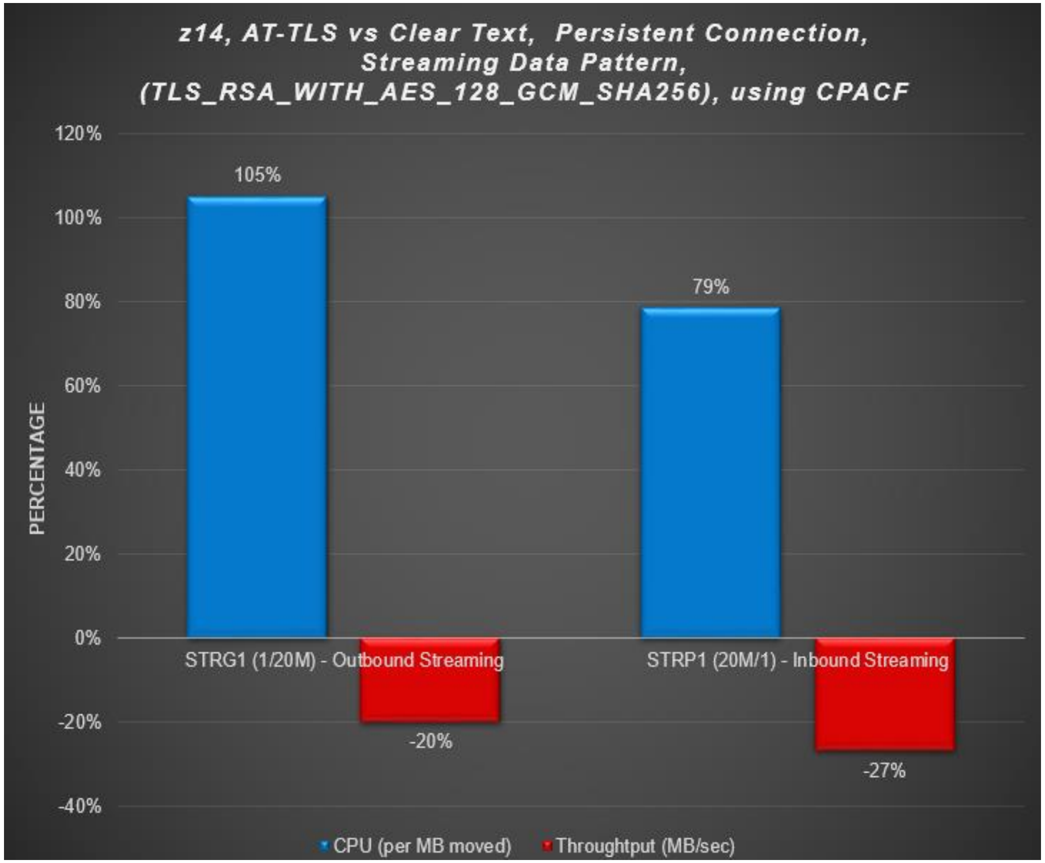
Notes:

- Micro-benchmark includes all z/OS host networking-related CPU overhead (up through the application socket layer)
- But the synthetic socket applications have *no* application logic
  - The networking related CPU cost equals entire application CPU cost
  - In real workloads, networking related CPU cost is fraction of overall application transaction cost (these benchmarks show the worst case scenario from a networking related CPU perspective)
- The TLS/SSL handshake occurs only at beginning of connections and is amortized across life of connection
- Main cost of TLS/SSL becomes encrypt/decrypt operations on bulk data flowing over connection

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3r Streaming Connections – CPU and Throughput Impact



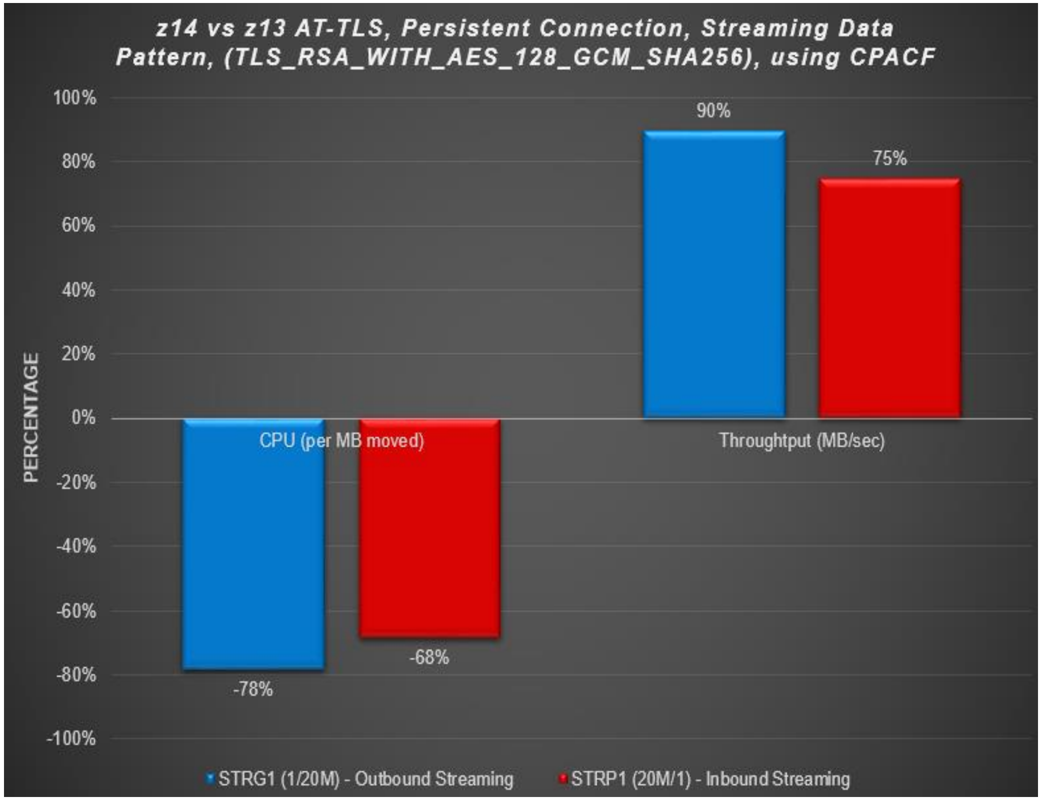
TLS/SSL processing for streaming TCP connections do not typically incur overhead of TLS/SSL handshakes

- TLS/SSL processing for streaming connections mainly consists of symmetric encryption/decryption operations that occur on z processors
- Cipher suites implemented on CPACF significantly improve performance – and z14 offers significant performance gains for selected cipher suites
- Observations:
  - Throughput impact is modest as number of network flows is not changed in streaming traffic patterns
  - CPU impact is largely associated with encrypting/decrypting large stream of data – in this benchmark, cost of encrypting data higher than cost to decrypt it

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3s Streaming Data - z14 vs z13 (CPACF Improvements)



Note- Left side two bar represents CPU benefits and right side two bar represents throughput benefits on z14 when compared to z13

Significant performance improvements for selected cipher suites in z14 CPACF

- Significant reduction in CPU cost – up to 78% lower CPU consumption per MB of data
- Significant increase in throughput (MB/sec) - up to 90%

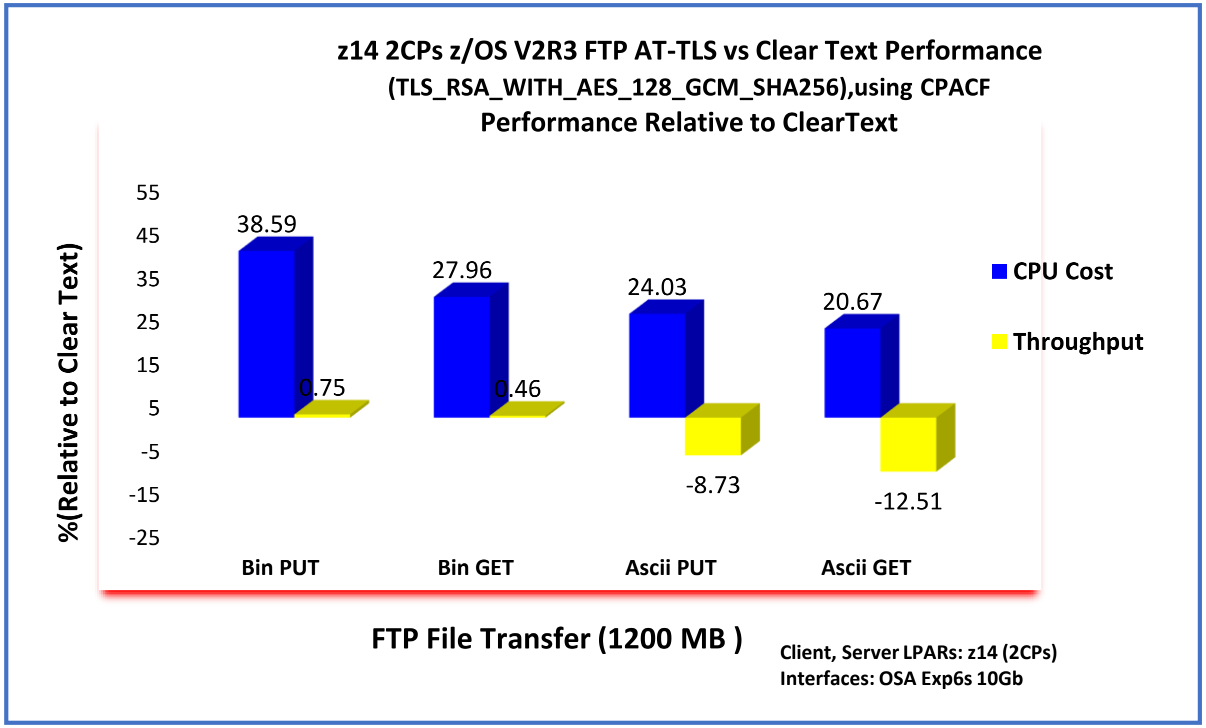
Recommendations:

- Evaluate performance benefits (CPU and latency) that z14 can offer your workloads compared to your current processors
  - Note: Performance benefits will be even more significant if you are migrating from older z processors
- Evaluate use of cipher suites that offer very compelling performance advantages on z14 CPACF (like AES GCM suites)

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3t FTP AT-TLS Performance – CPU and Throughput Impact



TLS/SSL processing for FTP TCP connections still incur overhead of TLS/SSL handshakes but it is very small percentage of overall cost.

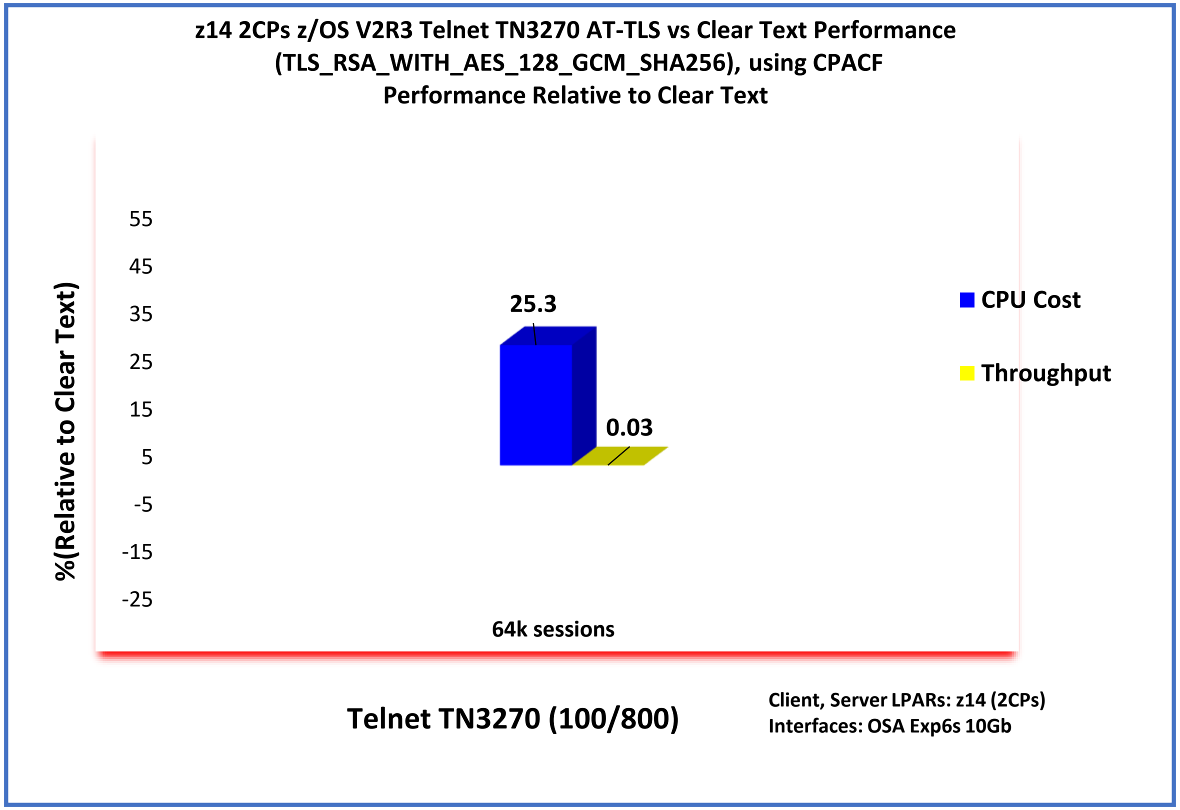
- TLS/SSL processing for FTP connections mainly consists of symmetric encryption/decryption operations that occur on z processors
- Cipher suites implemented on CPACF significantly improve performance

  - CPU Cost increased for AT-TLS was (20.67 to 38.59)% compared to Clear Text
  - With the AT-TLS there is no impact on throughput for binary transfer. For the AT-TLS ASCII transfer throughput was affected by (8.73 to 12.51)% when compared to Clear Text.

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3u Telnet AT-TLS Performance – CPU and Throughput Impact



TLS/SSL processing for Telnet Connections typically incur overhead of TLS/handshakes but observed minimum impact

- TLS/SSL processing for Telnet TN3270 steady state mainly consists of symmetric encryption/decryption operations that occur on z processors
- Cipher suites implemented on CPACF significantly improve performance
  - For 64K Telnet connection, CPU Cost increased by 25.3% for AT-TLS compared to Clear Text (Here the workload used was 100 byte request and 800 byte reply).

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

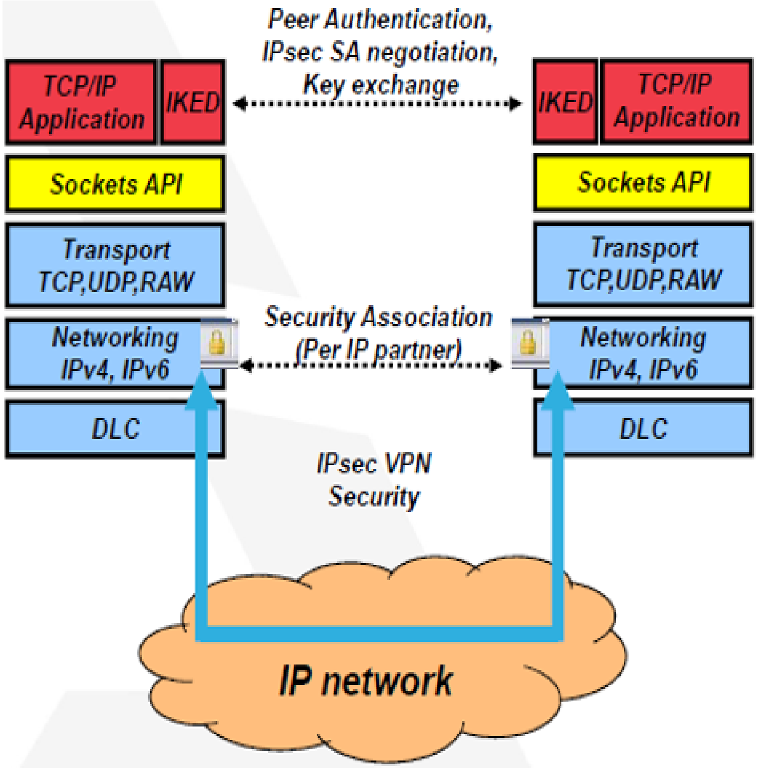
IP Security using IPSec

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3v IPSec Overview

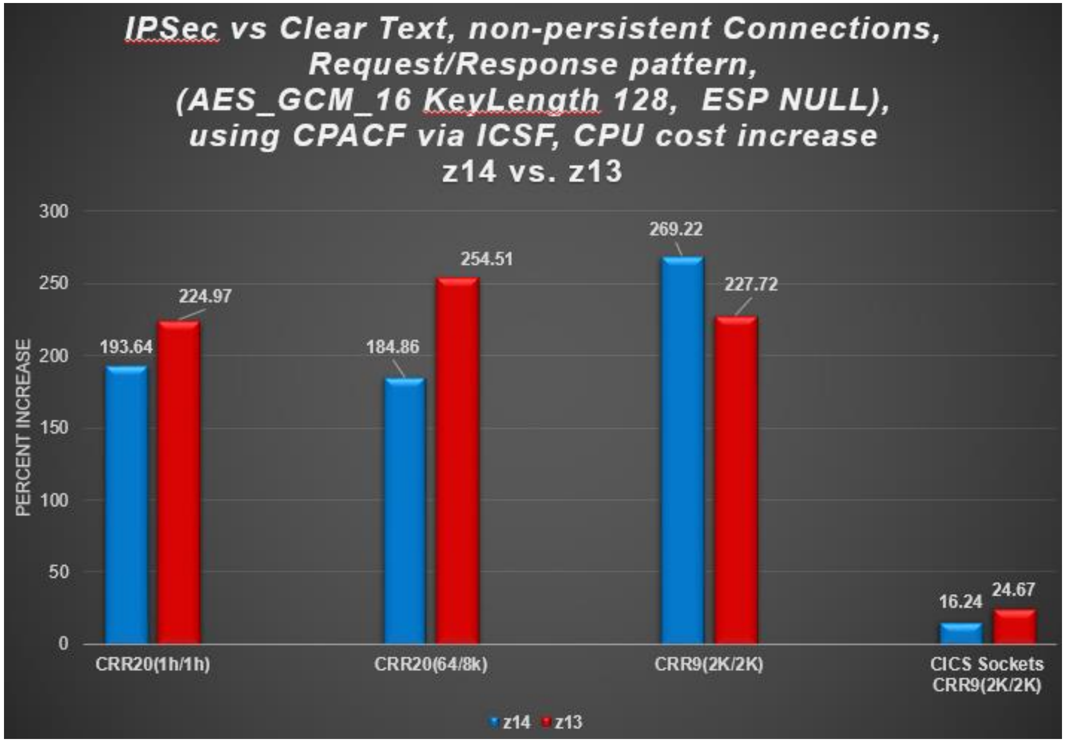
- Implemented at the IP (network) layer
  Completely transparent to application
  Supports all IP traffic, regardless of higher-layer protocols (suitable for Enterprise Extender)
- Node-to-Node protection via “Security Associations” (SAs)
  - –All traffic between nodes can use same security session
  - –Typically only negotiated when the VPN is established
- Data protection:
  - – Authentication Header (AH) provides data authentication and integrity protection
  - – Encapsulating Security Payload (ESP) provides data authentication, integrity protection, and encryption
- Management of crypto keys and security associations
  Dynamic through Internet Key Exchange (IKE)
  Manual
- Partner authentication via digital certificates using IKE protocol



V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3w Non-Persistent TCP Connections and IPSec – CPU impact



- Overhead for non-persistent TCP connections significantly lower than AT-TLS
  - No TLS/SSL handshake required on a per TCP connection basis

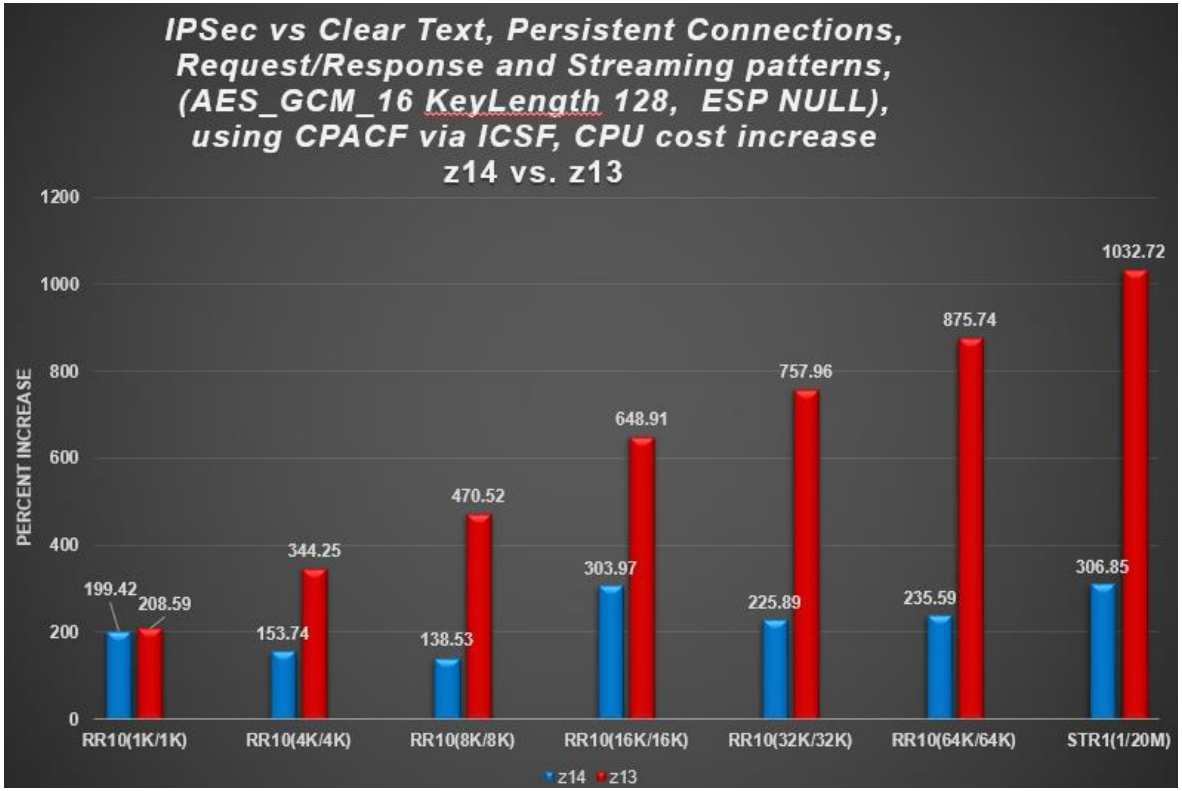
z14 offers lower CPU cost for most CRR patterns

- CPU cost of encryption for real workloads:
  - CICS Sockets CRR micro-benchmark shows significantly less CPU overhead than pure socket micro-benchmarks
  - CICS Sockets driving significantly more processing for scheduling and dispatching a CICS transaction for each connection
  - The CICS Sockets transaction has no application logic, it echoes back the data it received
  - Real CICS Socket workloads would experience a smaller percentage increase because they consume additional CPU cycles for application logic

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3x Request/Response and Streaming Data Patterns



z14 offers lower CPU cost for request/response and streaming patterns

- Significant savings vs z13 for larger payloads
- Even more significant savings if running on older z processors

AT-TLS is significantly more CPU efficient for securing these type of traffic patterns

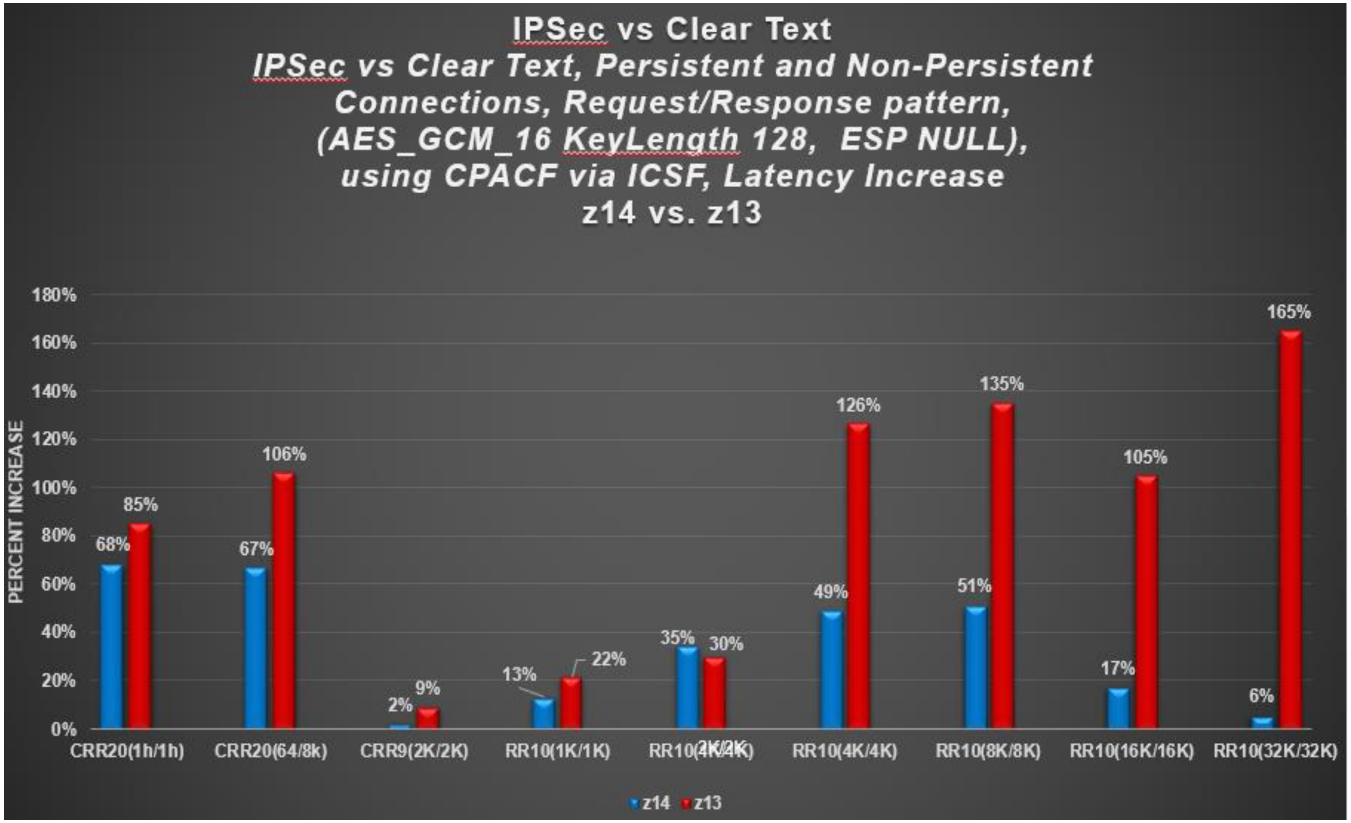
Contributing factors:

- Encrypting/Decrypting larger block of data
  - TLS/SSL supports up to 16K of data in a single segment – IPSec is based on packet size
  - TLS/SSL benefits from segmentation offload – IPSec cannot offload segmentation processing to OSA as each TCP packet requires encryption

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3y CRR and RR TCP Connections and IPSec – z14 vs z13 (Latency impact)



IPSec offers lower impact to latency for non-persistent connections

- No TLS/SSL handshake overhead and extra network roundtrip flows

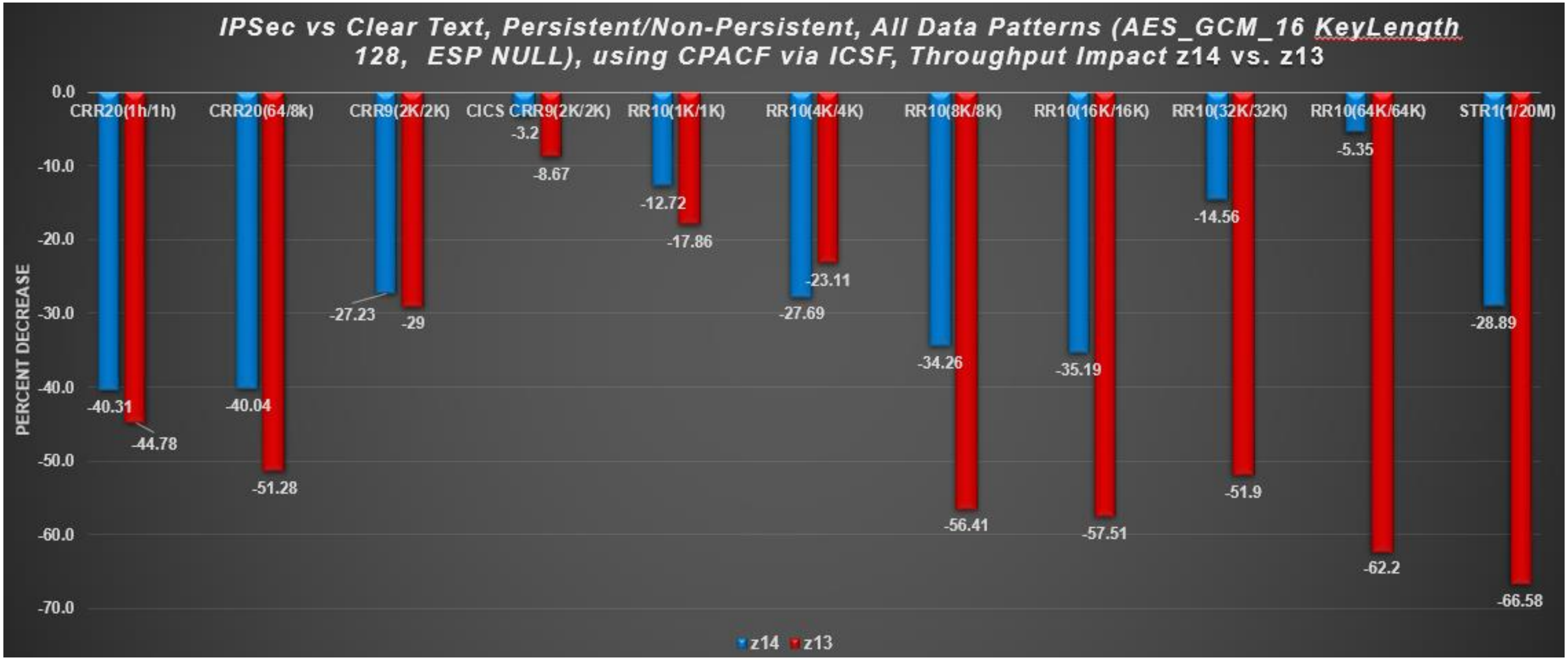
z14 provides significant latency improvements, especially for larger data due to CPACF improvements

V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3z IPSec vs Clear text – z14 vs z13 (Throughput impact)

Z14 offers significant performance (CPU, latency, and throughput) benefits – especially for larger payloads



V2R3 NETWORK SECURITY PERFORMANCE…

2.3 Network Security Deployment on z/OS

2.3za Best Practices

- Network Encryption can have a significant impact on CPU consumption, latency and throughput
  - *However, there are several ways to optimize and significantly reduce the impact of encryption and decryption*
    - Crypto Express provides significant acceleration and CPU reduction for TLS/SSL handshake processing – critical for short lived, non-persistent TCP connections
    - SSL/TLS session caching provides significant performance and CPU improvements for repeated, short-lived connections from the same clients
    - CPACF provides significant acceleration for encryption/decryption processing
    - z14 provides very significant improvements in performance with Crypto Express6S and CPACF for selected cipher suites and algorithms (like AES GCM suites)
      - z/OS V2R3 APAR number that we strongly recommend for GCM changes
        - ICSF : OA55958
        - SSL: OA55998

- It is important to perform capacity planning when enabling encryption/decryption for the first time
  - Including the correct number of Crypto Express6S features, available CPU, etc.
- Understanding a workload's communications and data patterns is key in this planning (persistent versus non-persistent connections, amount of data sent/received, request/response versus streaming)
  - The new 119 SMF records (subtypes 11,12) provided by zERT are a great source of information

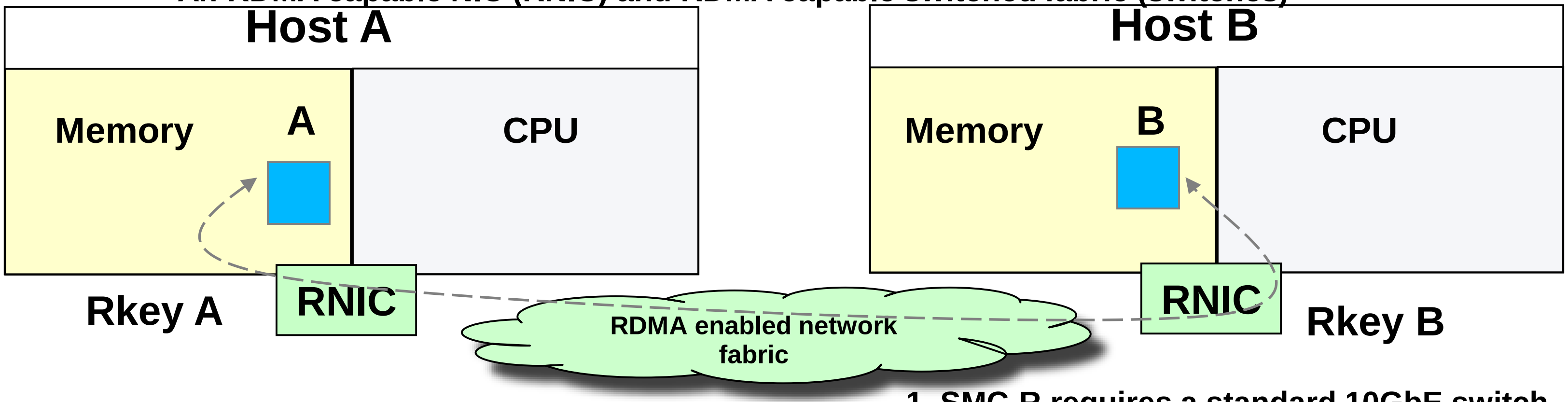
Shared Memory Communications Over RDMA (SMC-R)

Shared Memory Communications over RDMA (SMC-R)

3.1a RDMA (Remote Directory Memory Access) Technology Overview

Key attributes of RDMA

- **Enables a host to read or write directly from/to a remote host's memory *without* involving the remote host's CPU**
  - **By registering specific memory for RDMA partner use**
  - **Interrupts *still required* for notification (i.e. CPU cycles are not completely eliminated)**
- **Reduced networking stack overhead by using streamlined, low level, RMDA interfaces**
  - **Low level APIs such as uDAPL, MPI or RDMA verbs allow optimized exploitation**
    - ***For applications/middleware willing to exploit these interfaces***
- **Key requirements:**
  - **A reliable "lossless" network fabric (LAN for layer 2 data center network distance)**
  - **An RDMA capable NIC (RNIC) and RDMA capable switched fabric (switches)[1]**



[1] SMC-R requires a standard 10GbE switch

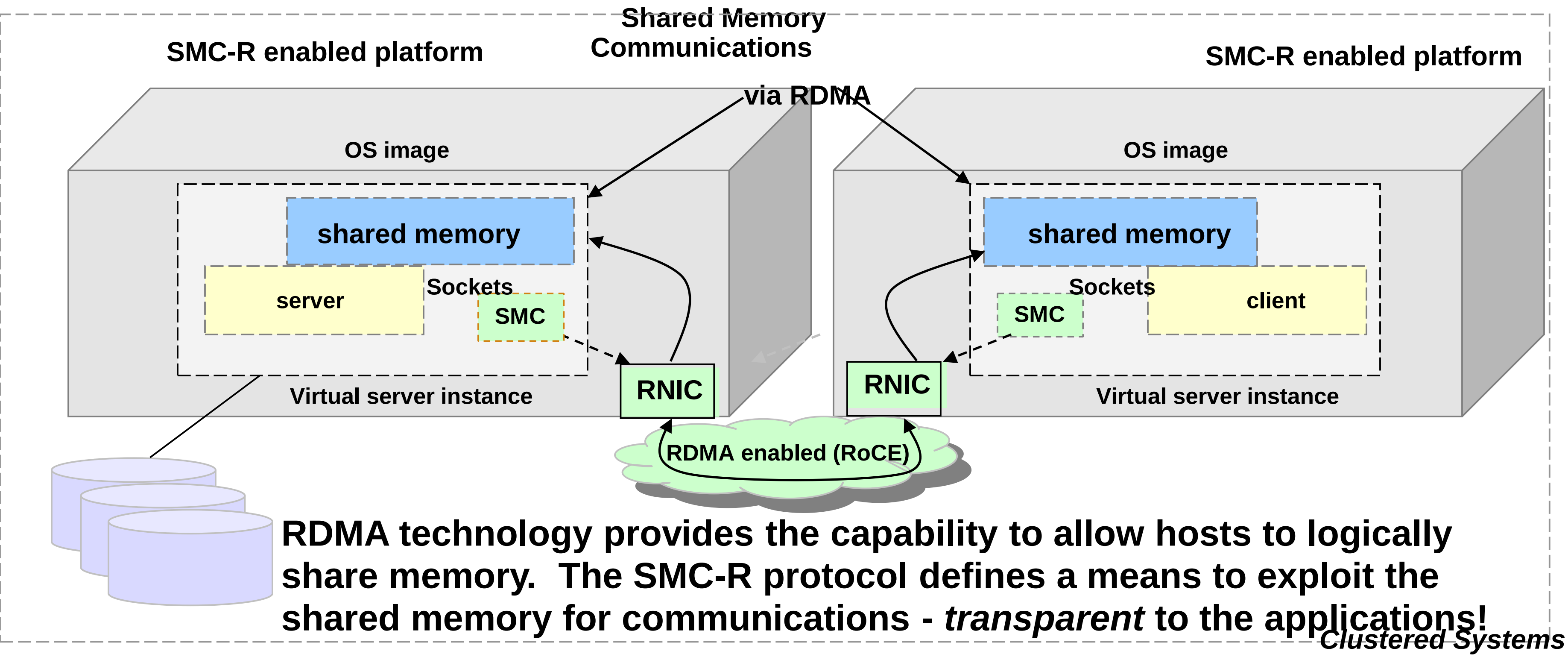
Shared Memory Communications over RDMA (SMC-R)….

3.1b RoCE - RDMA over Converged (Enhanced) Ethernet

- RDMA based technology has been available in the industry for many years – primarily based on Infiniband (IB)
  - IB requires a completely unique network eco system (unique hardware such as host adapters, switches, host application software, system management software/firmware, security controls, etc.)
  - IB is popular in the HPC (High Performance Computing) space
- RDMA technology is now available on Ethernet – RDMA over Converged Ethernet (RoCE)
  - RoCE uses existing Ethernet fabric but requires advanced Ethernet hardware (RDMA capable NICs and RoCE capable Ethernet switches)
  - ***RoCE is a game changer!***
    - ***RDMA technology becomes more affordable and prevalent in data center networks***
- Host software exploitation options fall into two general categories:
  - Native / direct application exploitation
    - Several variations, all involve deep level of expertise in RDMA and a new programming model
  - ***Transparent application exploitation (e.g. sockets based)***
    - ***Improve Time To Value by automatically exploiting RDMA/RoCE for sockets based TCP applications***

Shared Memory Communications over RDMA (SMC-R)….

3.1c Review: Shared Memory Communications over RDMA (SMC-R)



RDMA technology provides the capability to allow hosts to logically share memory. The SMC-R protocol defines a means to exploit the shared memory for communications - *transparent* to the applications!

SMC-R is an *open* sockets over RDMA protocol that provides transparent exploitation of RDMA (for TCP based applications) while preserving key functions and qualities of service from the TCP/IP ecosystem that enterprise level servers/network depend on!

IETF RFC for SMC-R: http://www.rfc-editor.org/rfc/rfc7609.txt

Shared Memory Communications over RDMA (SMC-R)….

3.1d New innovations available on zBC12, zEC12 …z14

NEW	NEW	ENHANCED	ENHANCED	NEW
Data Compression Acceleration	**High Speed Communication Fabric**	**Flash Technology Exploitation**	**Proactive Systems Health Analytics**	**Hybrid Computing Enhancements**
Reduce CP consumption, free up storage & speed cross platform data exchange	Optimize server to server networking with reduced latency and lower CPU overhead	Improve availability and performance during critical workload transitions, now with dynamic reconfiguration; Coupling Facility exploitation (SOD)	Increase availability by detecting unusual application or system behaviors for faster problem resolution before they disrupt business	x86 blade resource optimization; New alert & notification for blade virtual servers; Latest x86 OS support; Expanding futures roadmap
zEDC Express	*10GbE RoCE Express*	*IBM Flash Express*	*IBM zAware*	*zBX Mod 003; zManager Automate; Ensembl Availability Manager; DataPower Virtual appliance SoD*

Shared Memory Communications over RDMA (SMC-R) ….

3.1e Optimize server to server networking – transparently

“HiperSockets™ like”capability across systems

Network latency for z/OS TCP/IP based OLTP workloads ***reduced*** by up to ***85%*****

Shared Memory Communications (SMC-R):

Exploit RDMA over Converged Ethernet (RoCE) to deliver superior communications performance for TCP based applications

Typical Client Use Cases:

Help to reduce both latency and CPU resource consumption over traditional TCP/IP for communications across z/OS systems

Any z/OS TCP sockets based workload can seamlessly use SMC-R without requiring any application changes

NEW **z/OS V2.1 SMC-R**

NEW **z/VM 6.3 support for guests**

NEW **10GbE RoCE Express**

** Based on internal IBM benchmarks in a controlled environment of modeled z/OS TCP sockets-based workloads with request/response traffic patterns using SMC-R (10GbE RoCE Express feature) vs TCP/IP (10GbE OSA Express feature). The actual response times savings any user will experience will vary.

Shared Memory Communications over RDMA (SMC-R) ….

3.1f Use cases for SMC-R and 10GbE RoCE Express for z/OS to z/OS communications

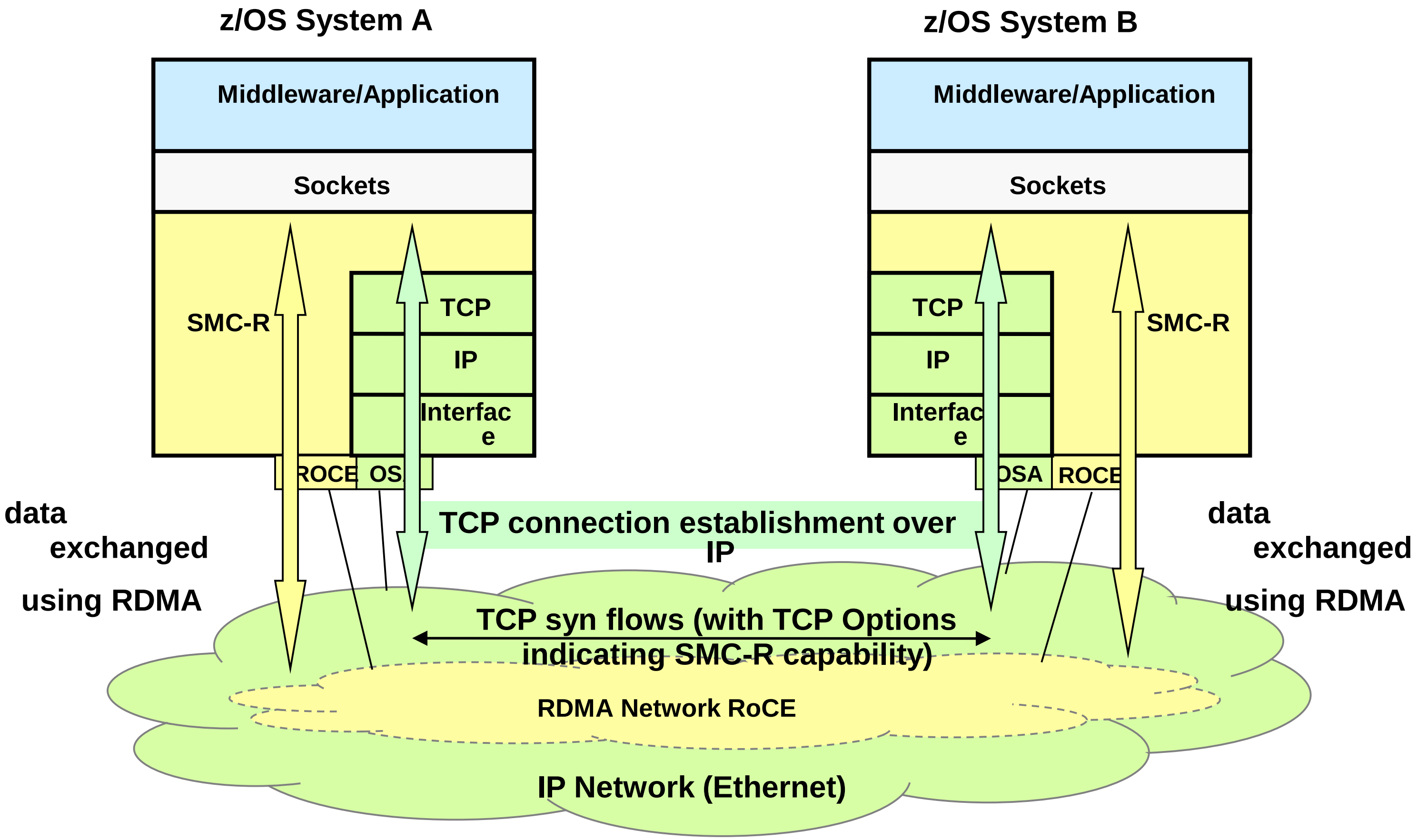
Use Cases

- Application servers such as the z/OS WebSphere Application Server communicating (via TCP based communications) with CICS, IMS or DB2 – particularly when the application is network intensive and transaction oriented
- Transactional workloads that exchange larger messages (e.g. web services such as WAS to DB2 or CICS) will see benefit.
- Streaming (or bulk) application workloads (e.g. FTP) communicating z/OS to z/OS TCP will see improvements in both CPU and throughput
- Applications that use z/OS to z/OS TCP based communications using Sysplex Distributor

Plus … *Transparent to application software – no changes required!*

Shared Memory Communications over RDMA (SMC-R) ….

3.1g Dynamic Transition from TCP to SMC-R



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

Shared Memory Communications over RDMA (SMC-R) ….

3.1h SMC-R Overview

- Shared Memory Communications over RDMA (SMC-R) is a protocol that allows *TCP sockets* applications to transparently exploit RDMA (RoCE)
- SMC-R is a “hybrid” solution that:
  - Uses TCP connection (3-way handshake) to establish SMC-R connection
  - Each TCP end point exchanges TCP options that indicate whether it supports the SMC-R protocol
  - SMC-R “rendezvous” (RDMA attributes) information is then exchanged within the TCP data stream (similar to SSL handshake)
  - Socket application data is exchanged via RDMA (write operations)
  - TCP connection remains active (controls SMC-R connection)
  - This model preserves many critical existing operational and network management features of TCP/IP

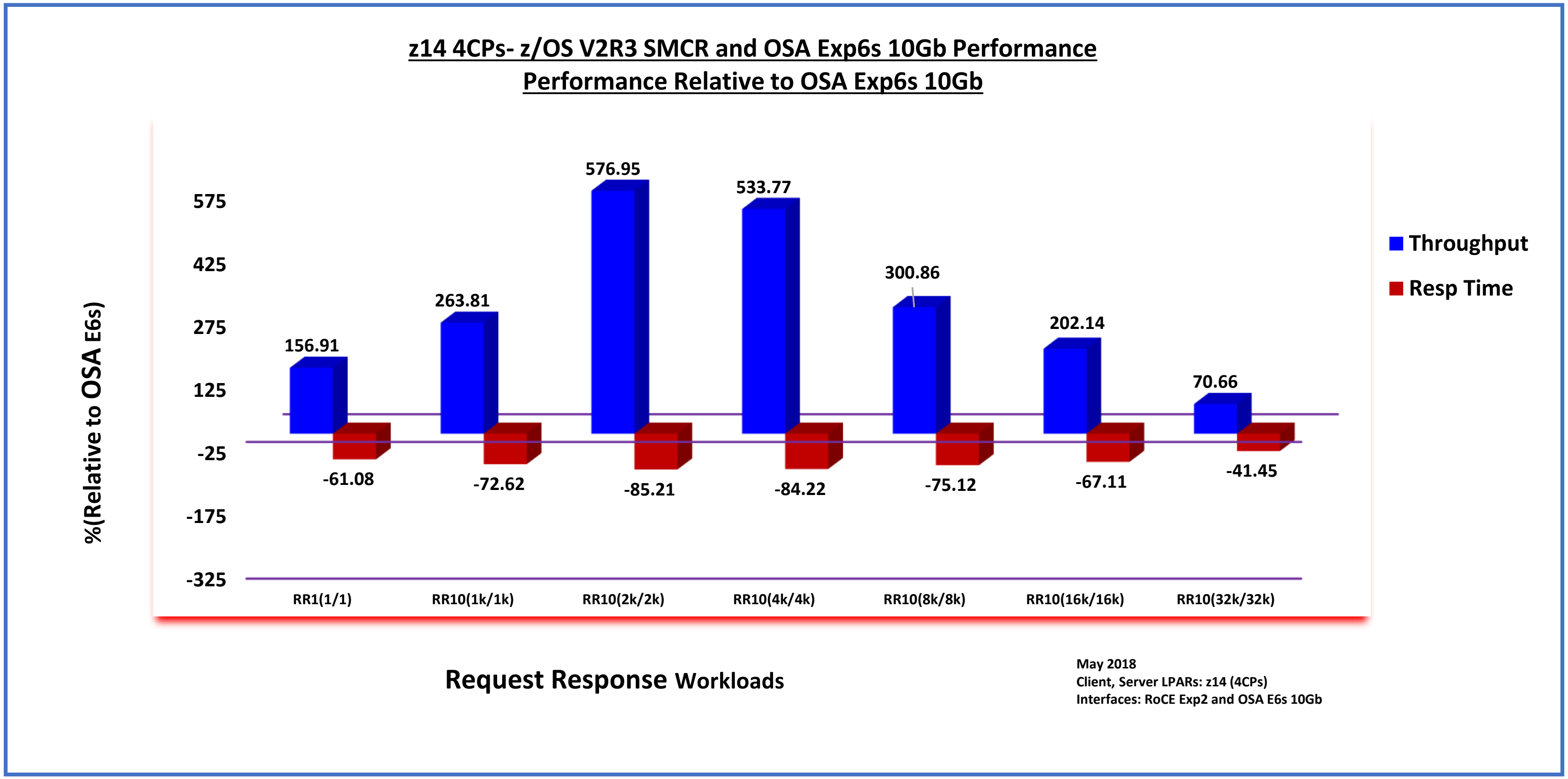
Shared Memory Communications over RDMA (SMC-R)….

3.1i Why a "Hybrid Protocol"? (Why TCP/IP + SMC-R?)

- The Hybrid model of SMC-R leverages key existing attributes:
  - Follows standard TCP/IP connection setup
  - Dynamically switches to RDMA (SMC-R)
  - TCP connection remains active (idle) and is used to control the SMC-R connection
  - Preserves critical operational and network management TCP/IP features such as:
    - Minimal (or zero) IP topology changes
    - Compatibility with TCP connection level load balancers (e.g Sysplex Distributor)
    - Preserves existing IP security model (e.g. IP filters, policy, VLANs, SSL etc.)
  - Minimal network admin / management changes
- *Significant reduction in Time to Value!*

Shared Memory Communications over RDMA (SMC-R) ….

3.1j z14 z/OS V2R3 SMC-R vs OSA Exp6s 10Gb Performance (Request Response Workloads)



Note- For OSA Exp6s used best performance practices SMCR provides (enables Large Send, jumbo frame and IWQ

SMCR provides Up to 6x the throughput and Up to 85% lower Response time compared to OSA Exp6s 10Gb.

Shared Memory Communications over RDMA (SMC-R) ….

3.1k SMC-R to OSA Summary Highlights

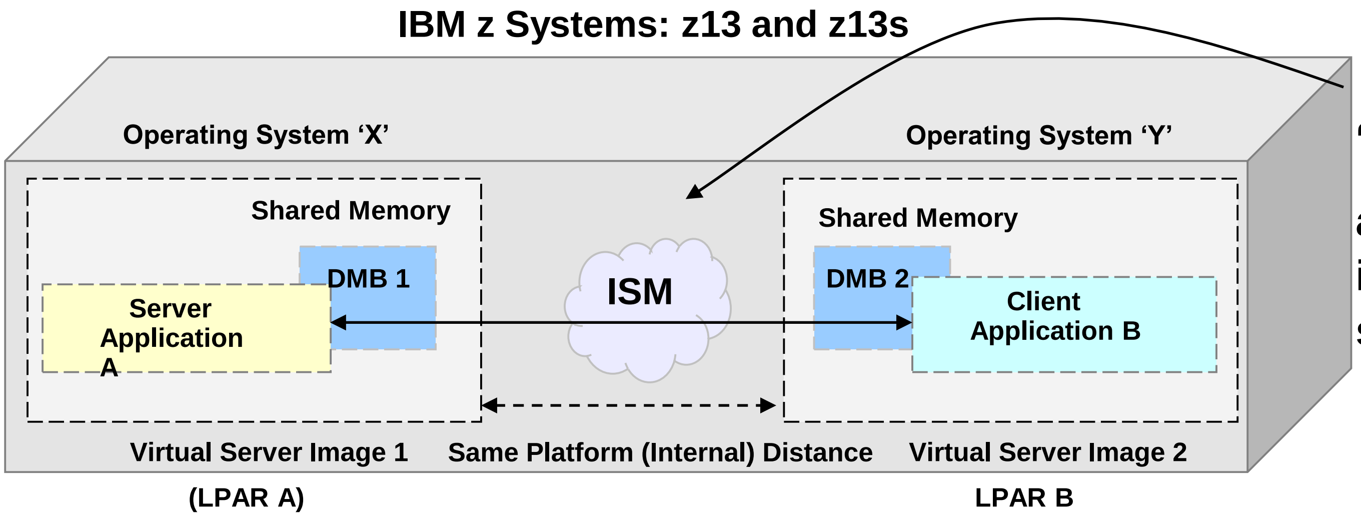
- **Request/Response Summary for Workloads with 1k/1k – 4k/4k Payloads:**
  - **Latency: Up to 85% reduction in latency**
  - **Throughput: Up to 576% (~6x) increase in throughput**

- **Request/Response Summary for Workloads with 8k/8k – 32k/32k Payloads:**
  - **Latency: Up to 75% reduction in latency**
  - **Throughput: Up to 300% (~3x) increase in throughput**

Shared Memory Communications – Direct Memory Access (SMC-D Introduction)

Shared Memory Communications-Direct Memory Access (SMC-D)

3.2a Shared Memory Communications-Direct Memory Access (SMC-D) over Internal Shared Memory (ISM)



"Shared Memory"

across unique OS instances within th 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 (over RoCE) 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.).

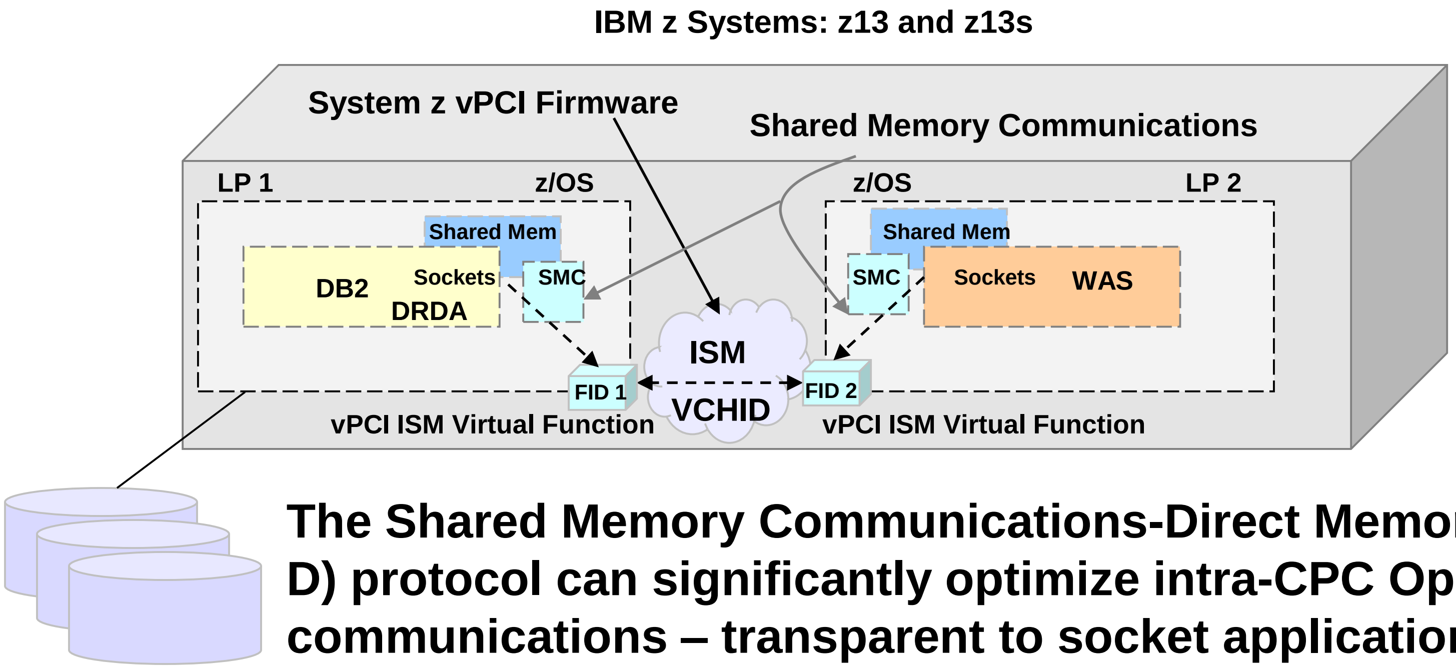
Note 1. The performance benefits of SMC-R (cross CPC) and HiperSockets (within CPC) are similar to each other.

SMC-D / ISM provides significantly improved performance benefits above both within the CPC.

Reference performance information: http://www-01.ibm.com/software/network/commserver/SMCR/

Shared Memory Communications – Direct Memory Access….

3.2b SMC-D over ISM: Internal Shared Memory vPCI Function with ISM VCHIDs



The Shared Memory Communications-Direct Memory Access (SMC-D) protocol can significantly optimize intra-CPC Operating Systems communications – transparent to socket applications!

•Tightly couples socket API communications / memory within the CPC.

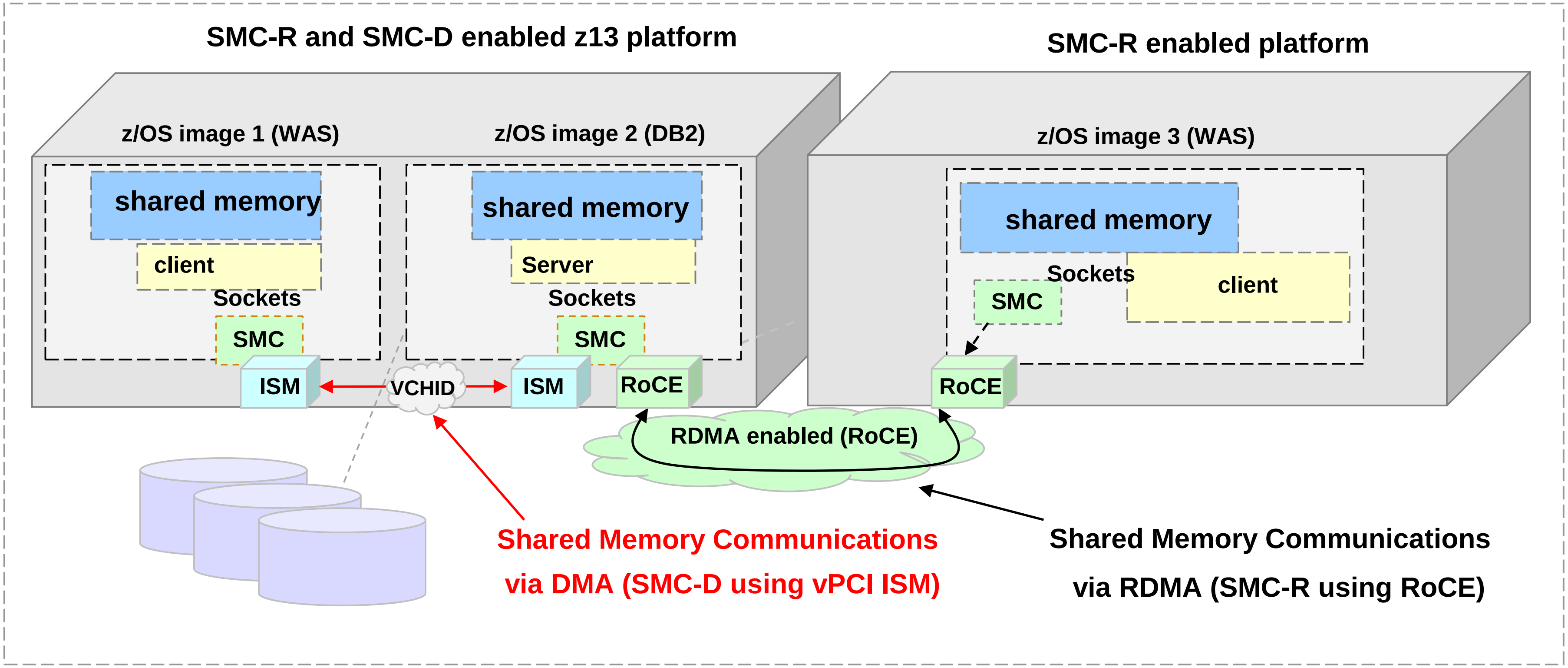
•Eliminates TCP/IP processing in the data path.

•ISM is a Z System firmware solution that leverages existing Operating System virtual memory PCI architecture without requiring any additional hardware.

Shared Memory Communications – Direct Memory Access (SMC-D) ….

3.2c Shared Memory Communications within the enterprise data center (RoCE) and within System z (ISM)

Clustered Systems: Example: Local and Remote access to DB2 from WAS (JDBC using DRDA)



Both forms of SMC can be used concurrently combining to provide a highly optimized solution.

Shared Memory Communications: via System z PCI architecture:

1. **RDMA (SMC-R for cross platforms via RoCE)**
2. **DMA (SMC-D for same CPC via ISM)**

Shared Memory Communications – Direct Memory Access (SMC-D)....

3.2d SMC-D Performance Benefits and Value (Performance Overview)

- The value of the next generation of highly optimized internal CPC communications is about **providing significantly improved network performance** using tightly coupled socket API communications / memory within the CPC **without additional hardware**
- Network **improvement attributes** are typically described as **latency, throughput, CPU cost and scalability.** Improvements in network performance can potentially improve (increase) application workload transaction rates while **reducing CPU cost**.
- The network latency characteristics provided by SMC-D are compelling:
  - Network latency is typically expressed as “network round trip time.” This latency attribute can translate to an improved overall application transaction rate for z/OS to z/OS workloads.
  - **Workloads that are network intensive and transaction oriented** (sometimes described as “request/response” workloads) -- that require multiple and even hundreds of network (“client/server”) flows to complete a single transaction **will realize the most benefit.**

Shared Memory Communications - Direct Memory Access (SMC-D) ....

3.2e Shared Memory Communications 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

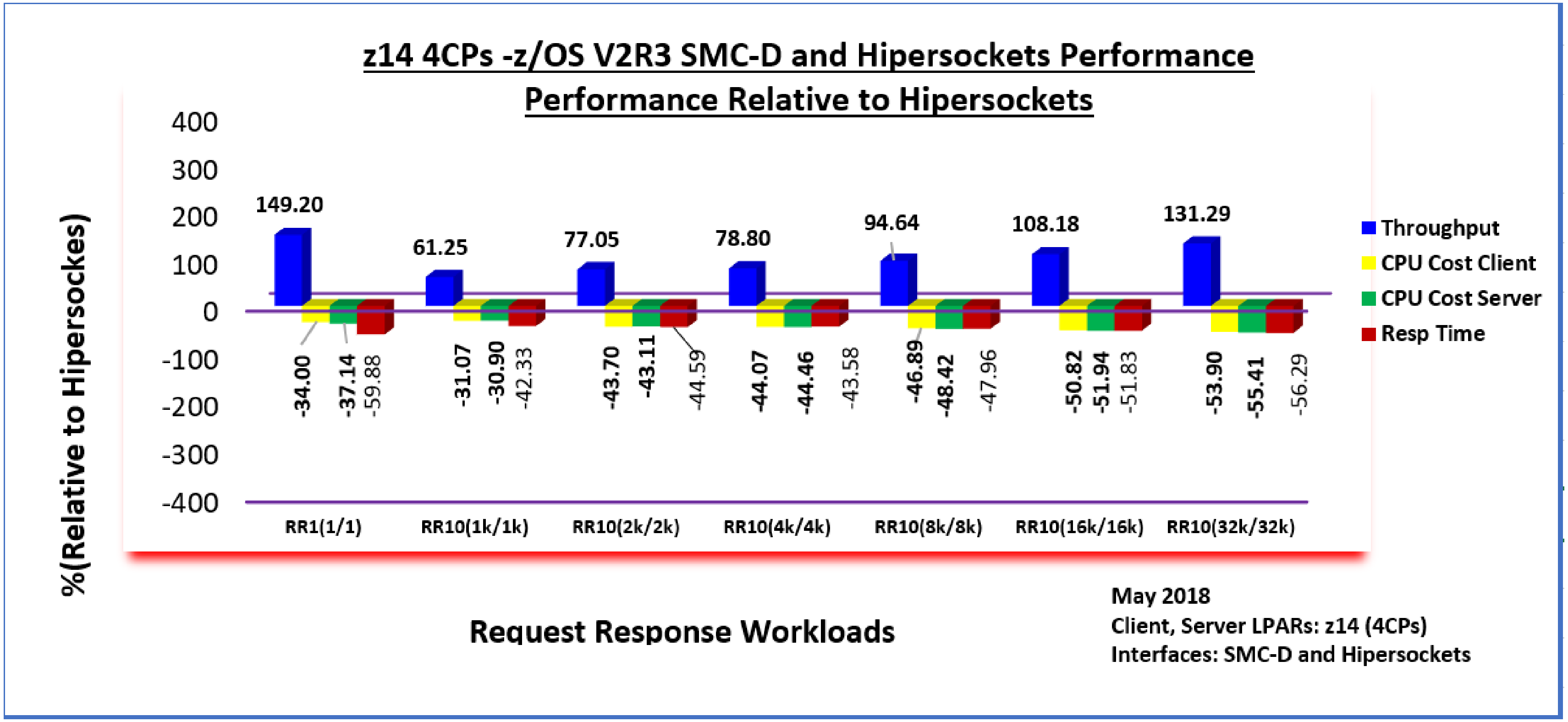
Shared Memory Communications - Direct Memory Access (SMC-D) ....

3.2f System z14 SMC-D Overall Performance Setup

- Performance results are based on IBM Internal micro benchmarks using standard tools used for z/OS release.
  - Environment :
    - Setup used: z14 4CPs Client, Server LPARs using same drawer with V2R2 Communications Server includes latest software and GA2 system firmware.
  - All Results Compared to HiperSockets using 16k/32k or 64k frame size.

Shared Memory Communications – Direct Memory Access (SMC-D) ….

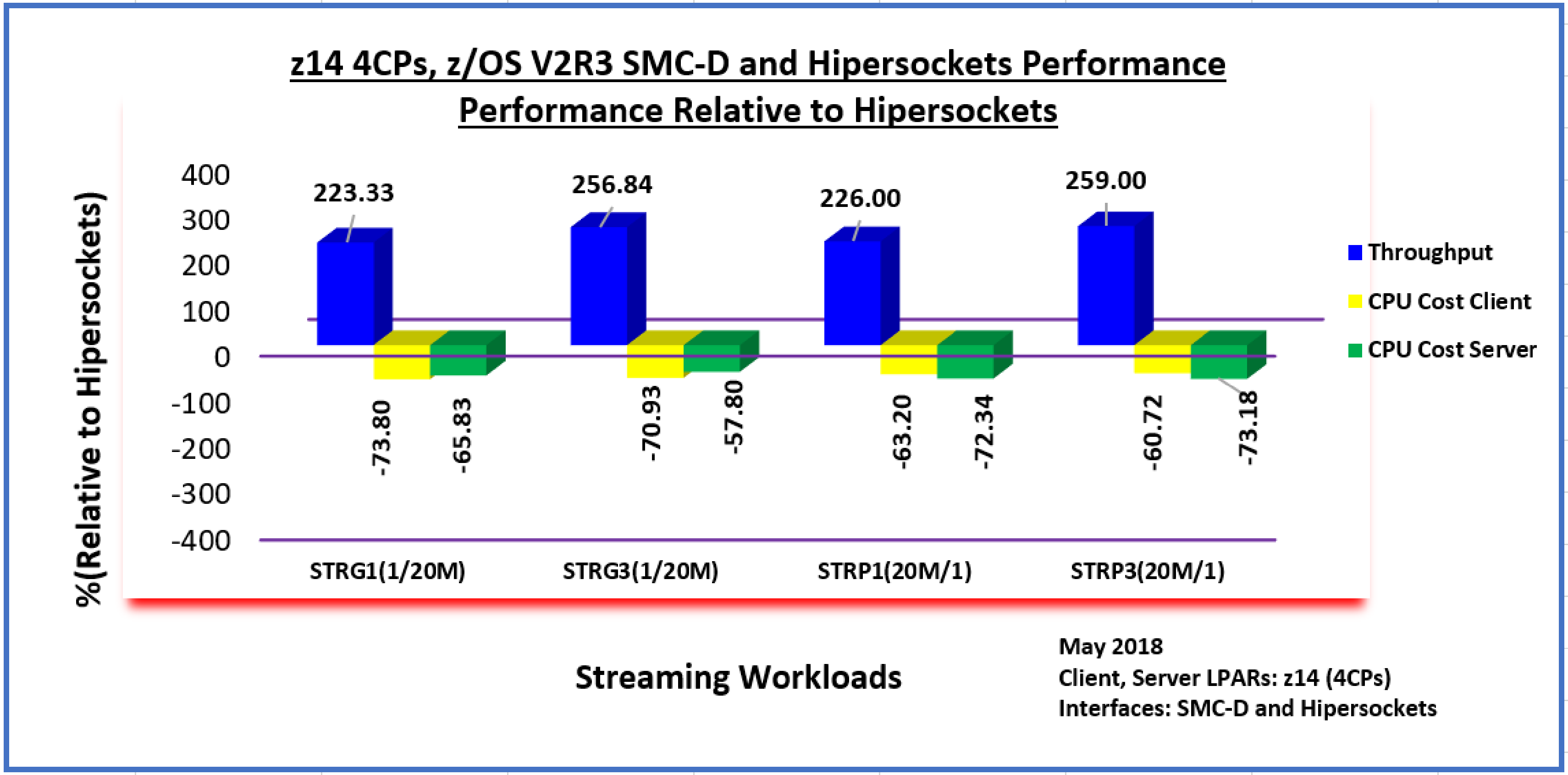
3.2g z14 z/OS V2R3: SMC-D vs HiperSockets Performance (Request Response Workloads)



- Up to 131% increase in throughput! See breakout summary charts
- SMC-D provide significantly lower CPU Cost compared to Hipersockets

Shared Memory Communications – Direct Memory Access (SMC-D) ….

3.2h z14 z/OS V2R3: SMC-D vs HiperSockets Performance (Streaming Workloads)



May 2018
Client, Server LPARs: z14 (4CPs)
Interfaces: SMC-D and Hipersockets

Up to 259% increase in throughput and Upto 73.80% lower CPU cost compared to HiperSockets.

Shared Memory Communications….

3.2i SMC-D / ISM to HiperSockets Summary Highlights

- **Request/Response Summary for Workloads with 1k/1k – 4k/4k Payloads:**
  - **Latency: Up to 44% reduction in latency**
  - **Throughput: Up to 79% increase in throughput**

- **Request/Response Summary for Workloads with 8k/8k – 32k/32k Payloads:**
  - **Latency: Up to 56% reduction in latency**
  - **Throughput: Up to 131% increase in throughput**
  - **CPU cost: Up to 55% reduction in network related CPU cost**

- **Streaming Workload**:
  - **Throughput: Up to 259% increase in throughput**
  - **CPU cost: Up to 73% reduction in network related CPU cost**

SMC Applicability Tool

SMC Applicability Tool

3.3a 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 ….

3.3b 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
  - 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)

SMC Applicability Tool ….

3.3c 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)*

SMC Applicability Tool ….

3.3d 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 Applicabiity Tool ….

3.3e 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).

SMC Applicabilty Tool ….

3.3f 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

```
                                          _INTERVAL 60_____
|---SMCATCFG____|__________________________|______________________________________________________>
                                          |_INTERVAL minutes |
>____________________________________________________________________________________________________|
   |                                                             |
   |                                                             |
   |                     ______________________________________  |
   |_IPADDR_____________V___ipv4_address_____________________|___|_
                          |_ipv4_address/num_mask_bits_|
                          |_ipv6_address_______________|
                          |_ipv6_address/prefix_length_|
```

SMC Applicability Tool ….

3.3g 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**

SMC Applicability Tool ….

3.3h 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,__ __________ __,__SMCAT,__ datasetname_______><
                   |_procname_|           |_,OFF__|
```

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

SMC Applicabilty Tool ….

3.3i 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

Reference Information

4.1a 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 2018 Summer Technical conference (St. Louis, MO)

  - **z/OS Communications Server: Technical Update, Part I and II (sessions 22818 and 22819)**
  - **IBM z/OS Communications Server Shared Memory Communications (SMC, Session 22803)**
  - **z/OS Communications Server Performance: Optimizing Your Network Encryption with z14 (session 22817)**
  - **TCPIP Security Controls on z/OS (session 22834)**

Reference Information …

4.1b 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.ibm.com/systems/z/os/zos/bkserv/v2r3pdf	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 ,Technotes & tips, White Papers, etc.)
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