



z/OS V2R1 Communications Server Technical Update

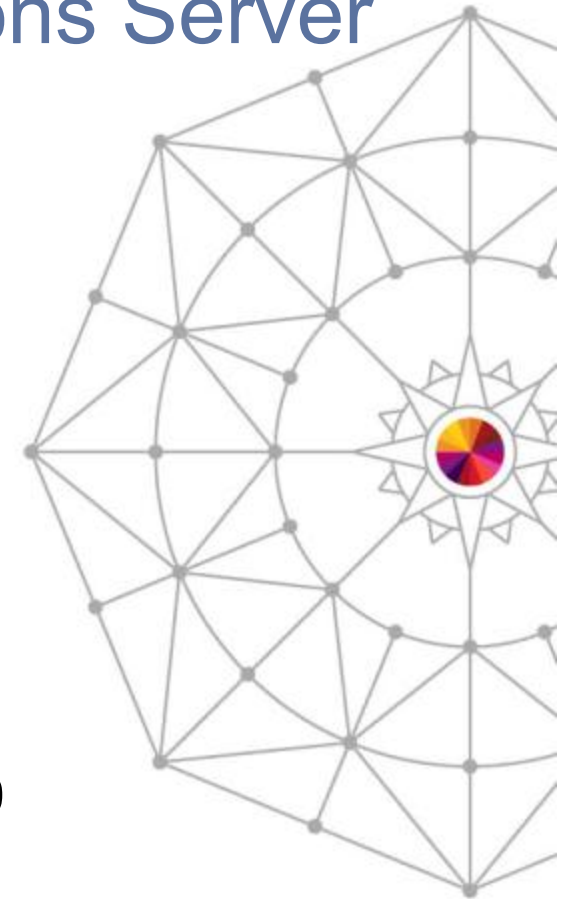
Gus Kassimis – kassimis@us.ibm.com

Sam Reynolds – samr@us.ibm.com

IBM Enterprise Networking Solutions

Monday, March 10, 2014 - 9:30 and 11:00

Sessions 14677 and 14678



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

* 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.

z/OS Communications Server Technical Update

Session number:	14677 and 14678
Date and time:	Monday, March 10, 2014 - 9:30 AM – 10:30 and 11:00 AM – 12:00 PM
Location:	Orange County Salon 1, Anaheim Marriott
Program:	Enterprise Data Center
Project:	Communications Server
Track:	Network Systems and z/OS Systems Programming
Classification:	Technical
Speaker:	Gus Kassimis, IBM Sam Reynolds, IBM
Abstract:	<p>z/OS Communication Server combines TCP/IP and VTAM support to better address the needs of today's complex networks. This two-part session provides an overview of features in the upcoming z/OS V2R1 Communications Server. Features to be discussed include:</p> <ul style="list-style-type: none"> • Enhanced fast path sockets • QDIO acceleration coexistence with IP filtering • A TCP/IP profile syntax check command • User control of ephemeral port ranges • A completely rewritten IBM Configuration Assistant for z/OS CS • Application-instance DVIPA affinity support • Enhanced security configuration options • ...and many more features related to areas such as network management, FTP, and Enterprise Extender

z/OS Communications Server and Social Media

- z/OS Communications Server has a blog, Facebook page, Twitter feed, and YouTube channel.
- We have been much more active via these channels in the last year.
- Follow us for announcements, hints and tips, screencasts on topics of interest, etc.



Find us on Facebook at
<http://www.facebook.com/IBMCommserver>



Follow us on Twitter at
http://www.twitter.com/IBM_Commserver



Read the z/OS Communications Server blog at
<http://tinyurl.com/zoscsblog>



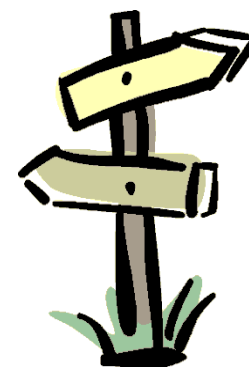
Visit the z/OS CS YouTube channel at
<http://www.youtube.com/user/zOSCommServer>



Agenda



- ❑ z/OS CS Requirements
- ❑ Shared Memory Communications over RDMA (SMC-R)
- ❑ Economics and Platform Efficiency
- ❑ Application / Middleware / Workload Enablement
- ❑ Statements of Direction
- ❑ Availability
- ❑ Simplification
- ❑ Security
- ❑ EE and SNA Enhancements
- ❑ Appendices



Disclaimer: All statements regarding IBM future direction or intent, including current product plans, are subject to change or withdrawal without notice and represent goals and objectives only. All information is provided for informational purposes only, on an "as is" basis, without warranty of any kind.

What is needed from z/OS networking in 2014 and beyond?

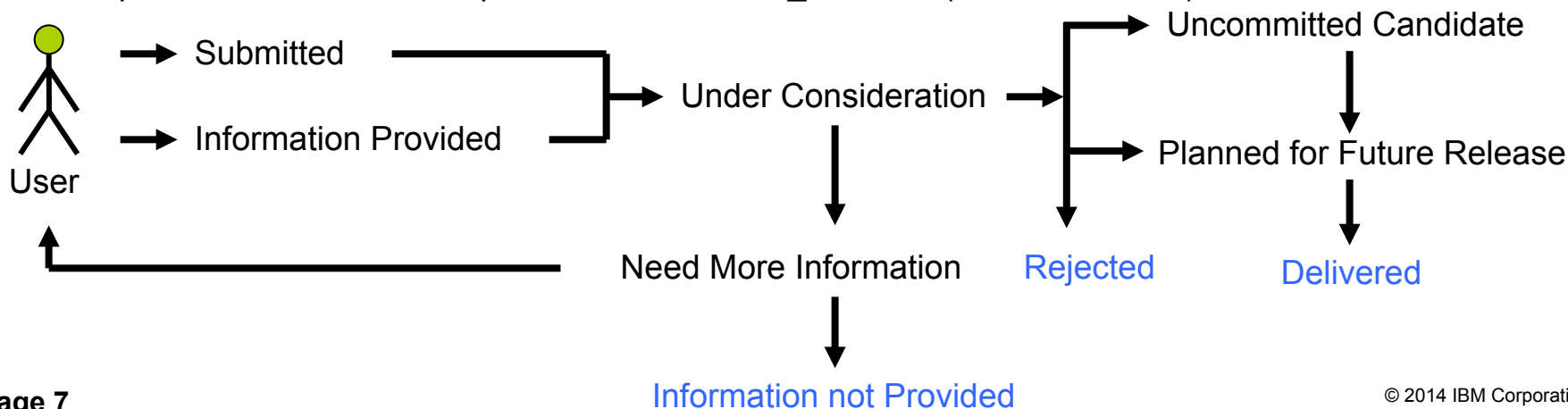


- **System z technology is expected to continue to evolve**
 - Networking software need to support new technologies
- **Access to System z system-level skills will continue to be an issue**
 - Retiring existing people, who grew up with system z
 - New people becoming responsible for the overall system z environment – including z/OS networking
 - Note: follow the IBM Academic Initiative
 - <https://www.ibm.com/developerworks/university/academicinitiative/>
- **Security will continue to be a hot topic**
 - Per customer survey, over 50% of network traffic will need encryption within the next few years
 - Trade organizations and governments continue to establish security and privacy compliance requirements that must be met
- **Price/performance requirements are high priority**
 - Continued demand for reduced cost in combination with increased performance and scalability on System z
- **Demand for increased “autonomic” system integration capabilities**
 - Continued demand for improved integration with other hardware and software platforms for more complex heterogeneous solutions
- **IANA has already run out of IPV4 addresses. Regional registries are also running out (APNIC and RIPE are both effectively exhausted, AFRINIC is getting close to exhaustion)**
 - IPv6 compliance (USGv6, IPv6-Ready, TAHI test suite, etc.)

RFE: New Requirements Process for z/OS Communications Server

▪ RFE (Request for Enhancement)

- New web-based interface for submitting requirements to some IBM products
- Unlike FITS, RFE is available to anyone that signs up for an IBM Universal ID.
- Can submit requirements, vote on requirements, watch and comment on requirements, and interact with IBM if more information is needed
- RFEs submitted against z/OS CS will initially be private, but most will be converted to public status after an initial screening.
- Even public RFEs do have certain fields that will always be private (only viewable by the submitter and IBM):
 - Customer name
 - Business justification
 - Attachments can be private (your choice at submission)
- Can keep up with requirements you are interested in via watchlists, bookmarks, and/or RSS feeds
- http://www.ibm.com/developerworks/rfe/?PROD_ID=498 (QR code above)



RFE: New Requirements Process for z/OS Communications Server ...

▪ Requirement Tips:

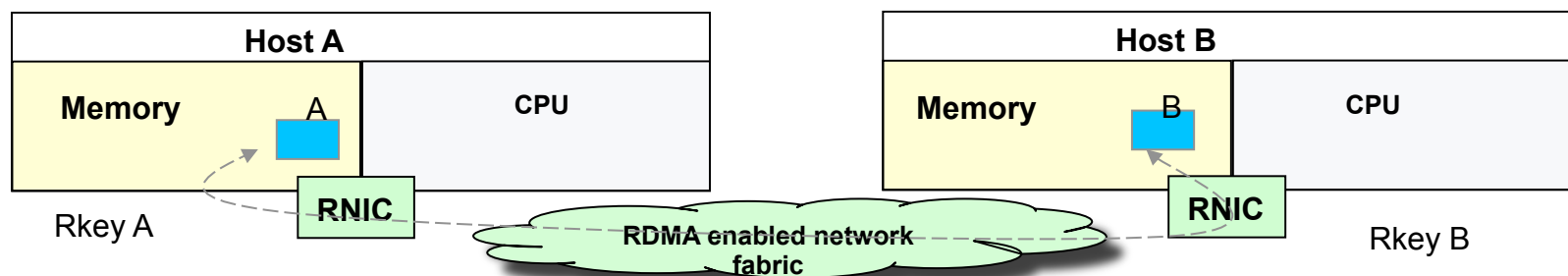
- Explain the problem you need solved, not just the requested solution
 - A request for a particular solution may not be feasible, or might take a very long time to deliver
 - By describing the problem to be addressed, we may be able to suggest alternatives that will be immediately beneficial, or a solution that we will be more likely able to implement in the reasonable future
 - Please understand that even in the best case, the delivery of a new function will likely be 2 to 3 years off due to our 2-year release cycle
- We try to disposition a requirement within a few weeks (typically marking it as an “uncommitted candidate” or closing it). During the time it is “Under Consideration” please monitor for our updates. We sometimes need to request more information, recommend alternate solutions, etc., before we can disposition it.
- If you see other z/OS Communications Server requirements that would be beneficial to your organization, please consider voting for them (as shown on the previous chart).

z/OS V2R1 Communications Server Technical Update

Shared Memory Communications over RDMA (SMC-R)



RDMA (Remote Direct Memory Access) Technology Overview



RDMA Technology

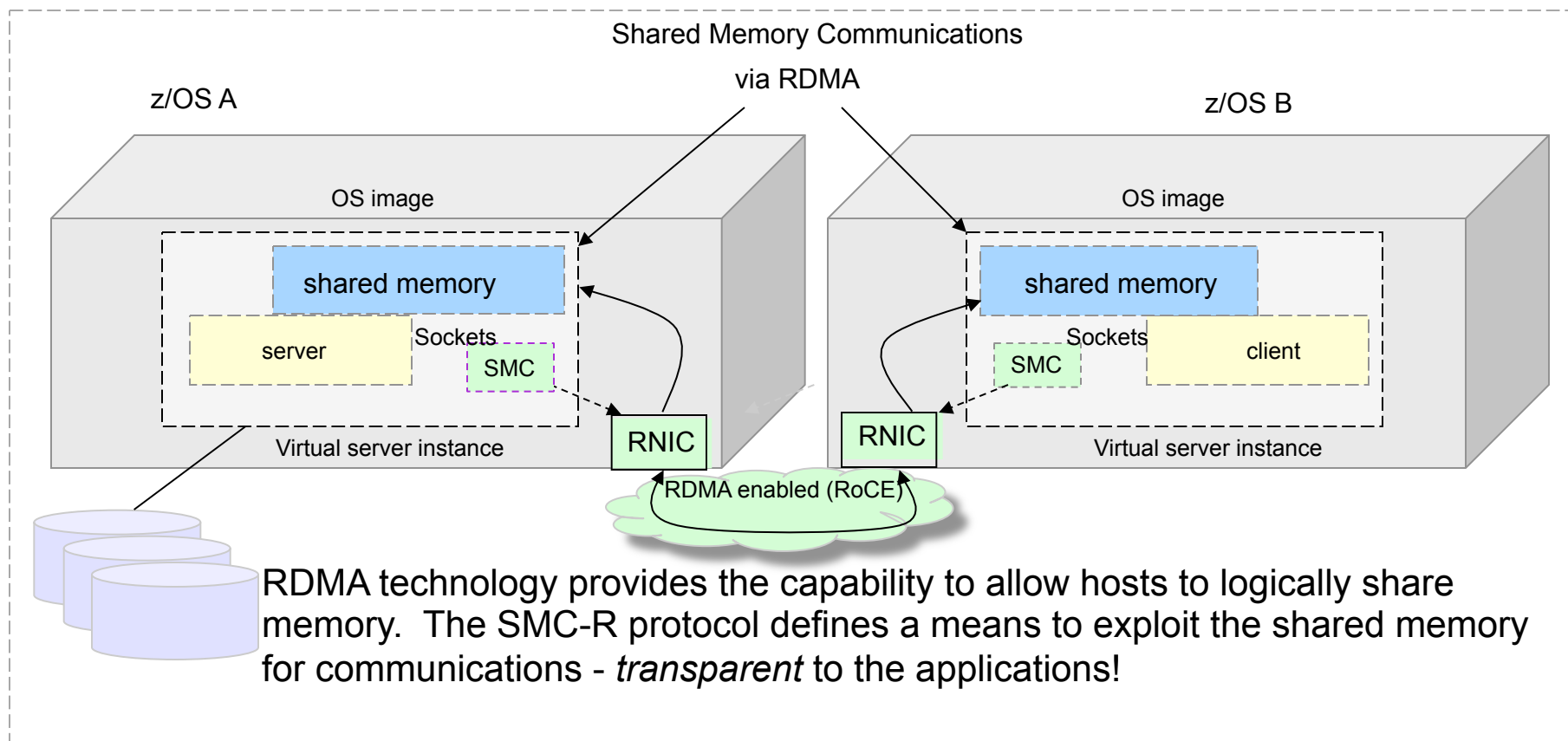
- Enables a host to read or write directly from/to a remote host's memory **without** involving the remote host's CPU
 - By registering specific memory for RDMA partner use
 - Interrupts **still required** for notification (i.e. CPU cycles are not completely eliminated)
- **Reduced** networking stack overhead (i.e. CPU overhead) and **lower network latency**
 - *Key requirements:* A reliable "lossless" network fabric
 - An RDMA capable NIC (RNIC) and RDMA capable switched fabric (switches)
- Options for exploiting this RDMA capability:
 - Native / direct application exploitation (requires middleware/application changes)
 - Transparent middleware/application exploitation (e.g. sockets-based, no code changes required)
 - **SMC-R provides this solution!**

Evolution of RDMA

- RDMA technology available for many years based on Infiniband (IB) technology, primarily in HPC space
 - Separate IB network and host adapters required (significant investment)
- RDMA technology now available on Ethernet: **RDMA over Converged Ethernet (RoCE)**
 - RoCE uses existing Ethernet fabric but requires advanced Ethernet hardware (RDMA capable NICs and CEE enabled Ethernet switches)
 - **Game changer: makes RDMA technology affordable for datacenter networks**

“Shared Memory Communications over RDMA” concepts

Clustered Systems



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!

This solution is referred to as *SMC-R* (Shared Memory Communications over RDMA). 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!

Draft IETF (Internet Engineering Task Force) RFC for SMC-R:

<http://tools.ietf.org/html/draft-fox-tcpm-shared-memory-rdma-03>

New innovations available on zBC12 and zEC12

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

14651: Introducing the new IBM zEnterprise BC12 and EC12 Updated Hardware: I/O Subsystem and Parallel Sysplex Coupling Function
Wednesday, March 12, 2014: 8:00 AM-9:00 AM Grand Ballroom Salon D (Anaheim Marriott) Speaker: Harv Emery (IBM)

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 **80%****



Shared Memory Communications (SMC-R):

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

Networking related CPU consumption for z/OS TCP/IP based workloads with streaming data patterns **reduced** by up to **60%** with a *network throughput* increase of up to **60%*****

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



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



**z/VM 6.3 support
for guests**



**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 and CPU savings any user will experience will vary.

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

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



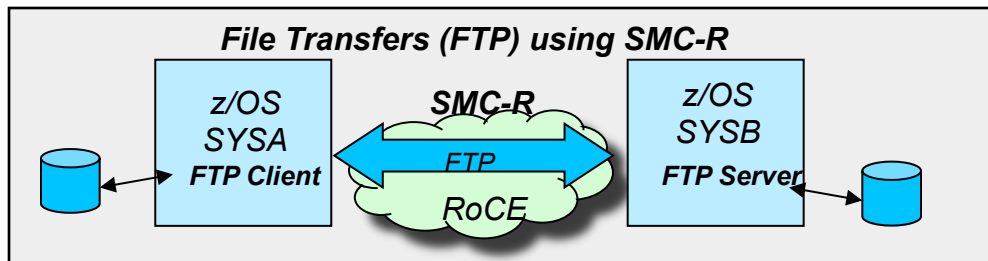
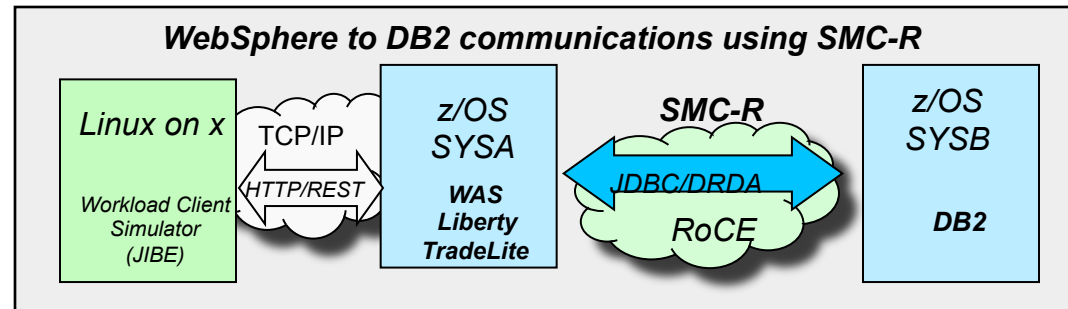
Use Cases

- 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!*

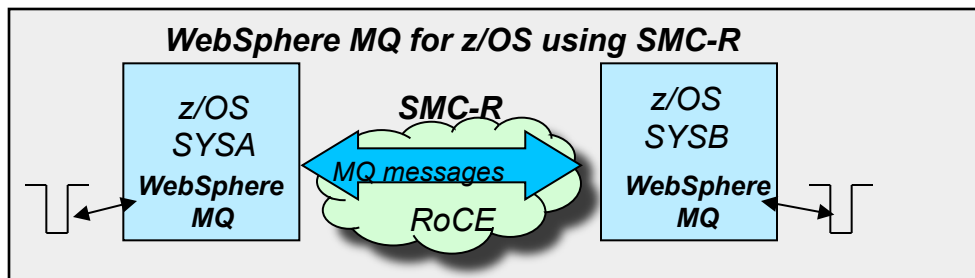
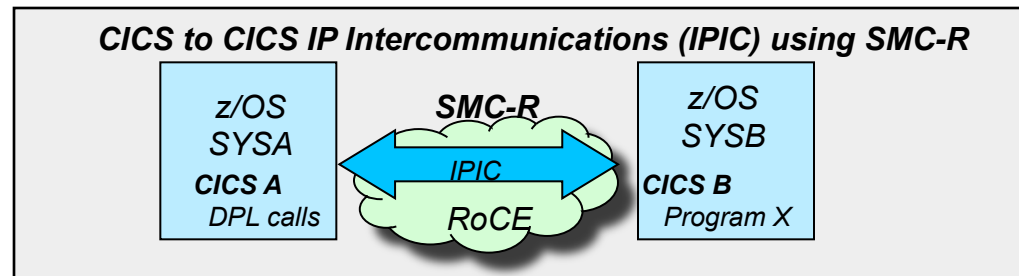
Impact of SMC-R on real z/OS workloads – early benchmark results

40% reduction in overall transaction response time for WebSphere Application Server v8.5 Liberty profile TradeLite workload accessing z/OS DB2 in another system measured in internal benchmarks *



Up to **50% CPU savings** for FTP binary file transfers across z/OS systems when using SMC-R vs standard TCP/IP **

Up to 48% reduction in response time and up to 10% CPU savings for CICS transactions using DPL (Distributed Program Link) to invoke programs in remote CICS regions in another z/OS system via CICS IP interconnectivity (IPIC) when using SMC-R vs standard TCP/IP ***



WebSphere MQ for z/OS **realizes up to 200% increase in messages per second** it can deliver across z/OS systems when using SMC-R vs standard TCP/IP ****

Note: Based on IBM internal benchmarks of modeled workloads in a controlled environment. The actual performance gains that users will experience may vary based on the user workload and configuration.

Performance benchmarks details and disclaimers

* Based on projections and measurements completed in a controlled environment. Results may vary by customer based on individual workload, configuration and software levels.

** Based on internal IBM benchmarks in a controlled environment using z/OS V2R1 Communications Server FTP client and FTP server, transferring a 1.2GB binary file using SMC-R (10GbE RoCE Express feature) vs standard TCP/IP (10GbE OSA Express4 feature). The actual CPU savings any user will experience may vary.

*** Based on internal IBM benchmarks using a modeled CICS workload driving a CICS transaction that performs 5 DPL calls to a CICS region on a remote z/OS system, using 32K input/output containers. Response times and CPU savings measured on z/OS system initiating the DPL calls. The actual response times and CPU savings any user will experience will vary.

**** Based on internal IBM benchmarks using a modeled WebSphere MQ for z/OS workload driving non-persistent messages across z/OS systems in a request/response pattern. The benchmarks included various data sizes and number of channel pairs. The actual throughput and CPU savings users will experience may vary based on the user workload and configuration.

Additional sessions related to SMC-R

14680: z/OS V2R1 CS: Shared Memory Communications - RDMA (SMC-R), Part 1
Tuesday, March 11, 2014: 9:30 AM-10:30 PM
Grand Ballroom Salon G (Anaheim Marriott)
Speaker: [Gus Kassimis](#) (IBM Corporation)

14681: z/OS V2R1 CS: Shared Memory Communications - RDMA (SMC-R), Part 2
Tuesday, March 11, 2014: 11:00 AM-12:00 PM
Grand Ballroom Salon G (Anaheim Marriott)
Speaker: [Dave Herr](#) (IBM Corporation)

15193: z/OS V2R1 Communications Server Performance Update
Wednesday, March 12, 2014: 11:00 AM-12:00 PM
Grand Ballroom Salon A (Anaheim Marriott)
Speaker: [Dave Herr](#) (IBM Corporation)

SMC-R References

- SMC-R Overview
 - <https://share.confex.com/share/121/webprogram/Session13627.html>
 - Overview with audio (youtube):
 - http://www.youtube.com/watch?v=8_5JviApQXw
- SMC-R Implementation:
 - <https://share.confex.com/share/121/webprogram/Session13628.html>
 - With audio (youtube):
 - <https://www.youtube.com/watch?v=TN0eS-l1FoE>
- Shared Memory Communications over RDMA: Performance Considerations (White Paper)
 - <http://www-01.ibm.com/support/docview.wss?uid=swg27041273>
- Performance information:
 - <https://share.confex.com/share/121/webprogram/Session13633.html>
- FAQ:
 - <https://www-03.ibm.com/support/techdocs/atmastr.nsf/WebIndex/FQ131485>
- RFC:
 - <http://tools.ietf.org/html/draft-fox-tcpm-shared-memory-rdma-03>
- SMC-R and Security Considerations White Paper:
 - <http://w3.ibm.com/sales/support/ShowDoc.wss?docid=ZSW03255USEN>

z/OS V2R1 Communications Server Technical Update

Economics and platform efficiency

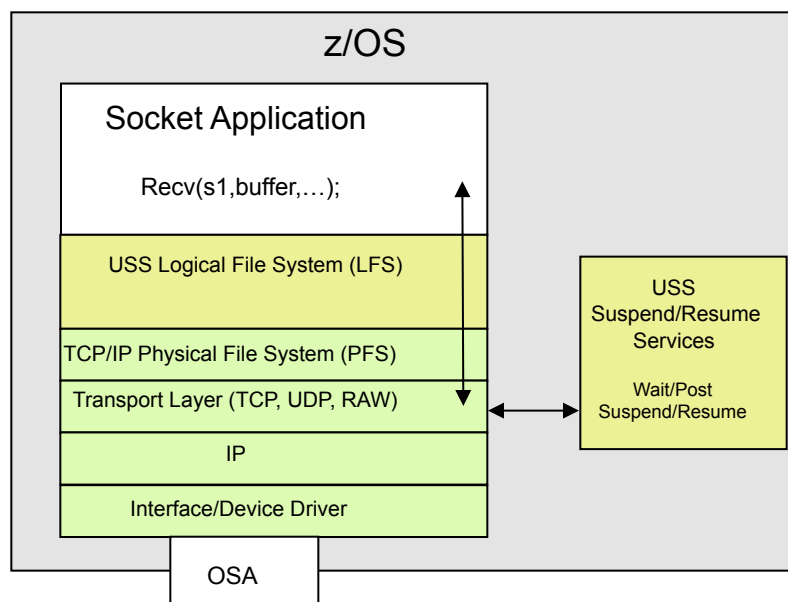


Removal of BIND DNS from z/OS

- The z/OS BIND implementation was not current on standards, and has been removed from z/OS V2R1 CS
 - There are RFCs to address BIND-related issues, and those RFCs are incompatible with the z/OS version of BIND
- Most z/OS customers run the DNS function on other platforms, since there is no differentiating advantage to running it on z/OS
- z/OS V1R11's resolver caching function eliminated the need for caching-only name servers, which was the most common use of DNS on z/OS
- We have previously issued Statements of Direction (SOD) for removal of BIND DNS from z/OS
 - V1R11 preview RFA indicated BIND would be removed from a "future" z/OS release
 - V1R13 preview RFA announced that V1R13 was the last release that would ship with BIND DNS
- We will continue to ship the DNS utility programs (nslookup, dig, etc.)

Enhanced Fast Path socket support

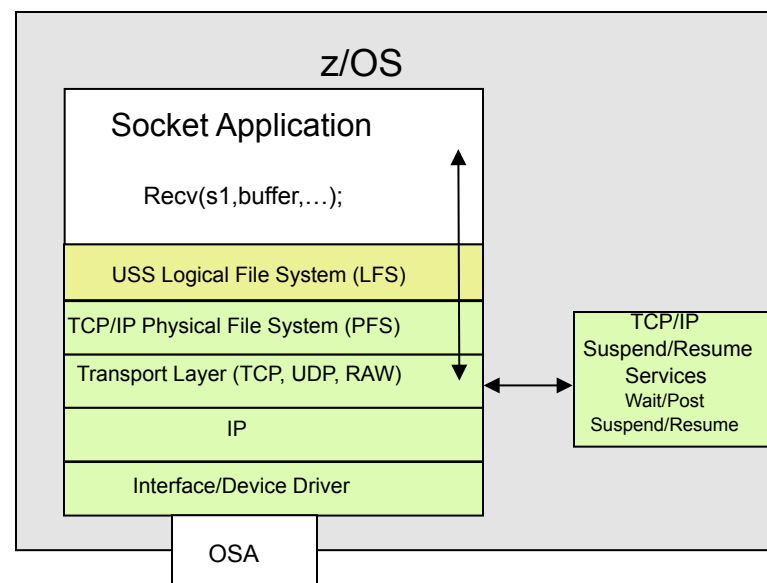
TCP/IP sockets (normal path)



Key attributes:

- Full function support for sockets, including support for Unix signals, POSIX compliance
- When TCP/IP needs to suspend a thread waiting for network flows, USS suspend/resume services are invoked.
 - These services require a space switch into OMVS address space
 - Allows thread to be woken up when data arrives or when a signal needs to be delivered

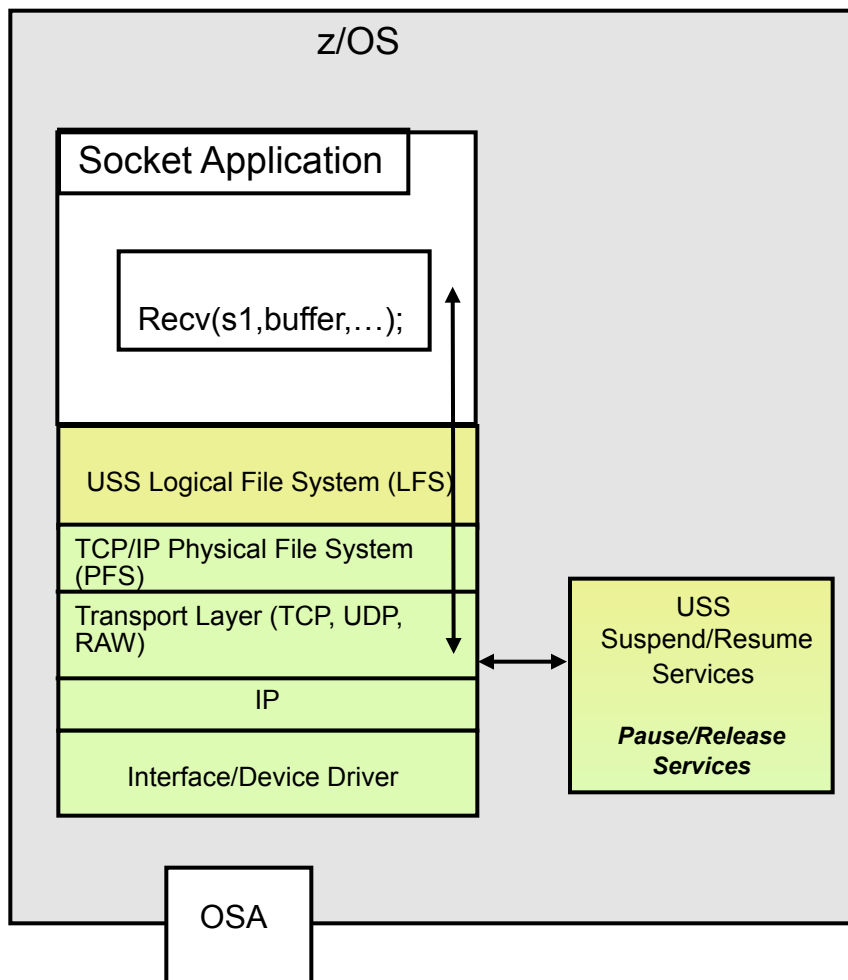
TCP/IP fast path sockets (existing support)



Key attributes:

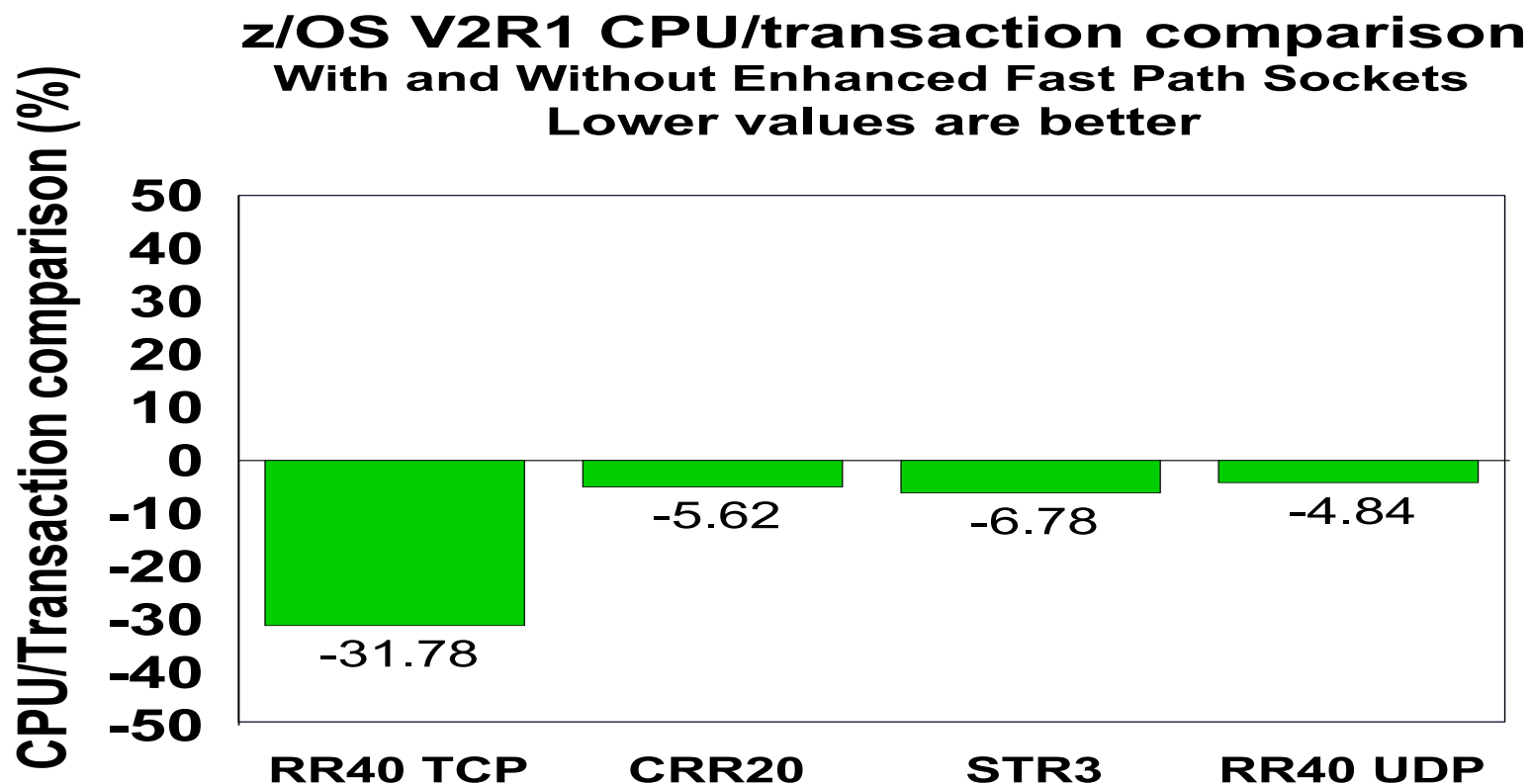
- Streamlined path through USS LFS (reduced path length) for selected socket APIs (send/rcv)
- When TCP/IP needs to suspend a thread waiting for network flows, TCP/IP performs the wait/post or suspend/resume inline using its own services
- Not POSIX compliant, no support for Unix signals (other than SIGTERM), no DBX support
- But significant reduction in path length for request/response workloads
- Must be explicitly enabled via socket API options (lcc#FastPath IOCTL) or via Environment variables (`_BPXK_INET_FASTPATH`)

Enhanced Fast Path socket support (new in z/OS V2R1)



- Provide fast path sockets like performance for *all* sockets applications
 - Without requiring explicit enablement by the application or the administrator
 - With POSIX compliance, signals support and DBX support
 - Valid of all socket APIs (except Pascal Socket API)
- Streamlined path through USS LFS for *recv/send* set of APIs
 - Enabled for Recv, recvmsg, recvfrom, send, sendmsg, sendto)
 - But not enabled for read, readv, write,writev, etc.
- New efficient Unix Services for suspend/resume processing
 - Runs as an extension of TCP/IP stack
 - No need to space switch into OMVS address space
 - Exploits MVS Pause/Release services
 - Optimized path length
 - Minimizes local lock contention for address spaces with large number of threads performing socket calls

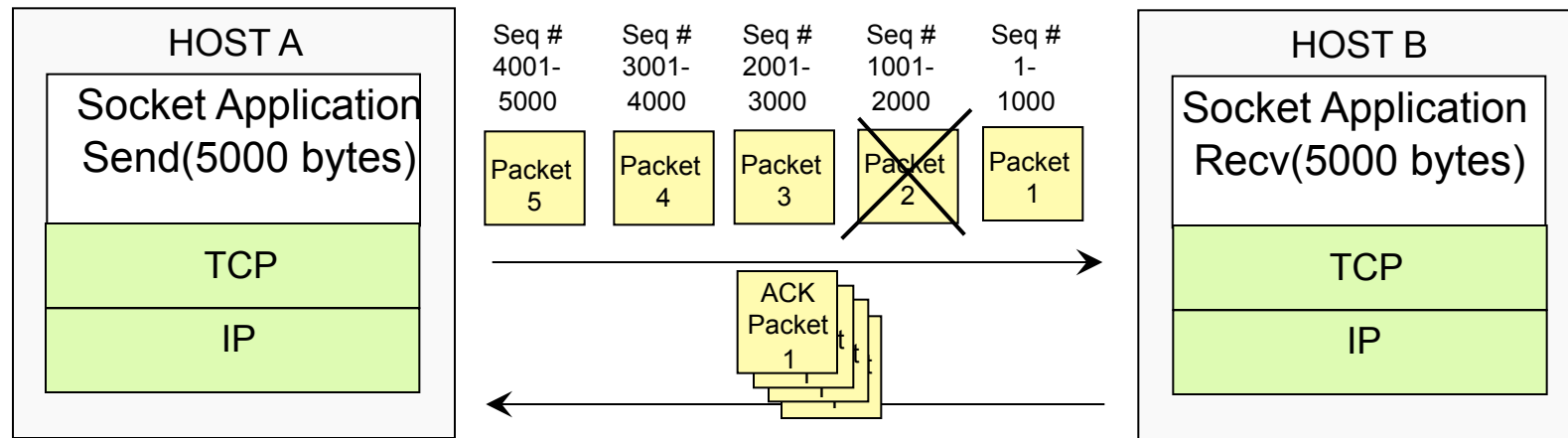
Enhanced Fast Path sockets – Performance Benchmarks



- / Request-Response, Connect-Request-Response, and Streams workloads
- / RR40 TCP: 40 sessions, TCP, 100 / 800 bytes; CRR20: 20 sessions, 64 / 8 KB; STR3: 3 sessions, 1 / 20 MB; RR40 UDP: 40 sessions, UDP, 100 / 800 bytes
- / All transactions are memory to memory (no DASD used)
- / Hardware: zEC12 using OSA-E4 (10 GbE) INBPERF DYNAMIC used; Software: z/OS V2R1

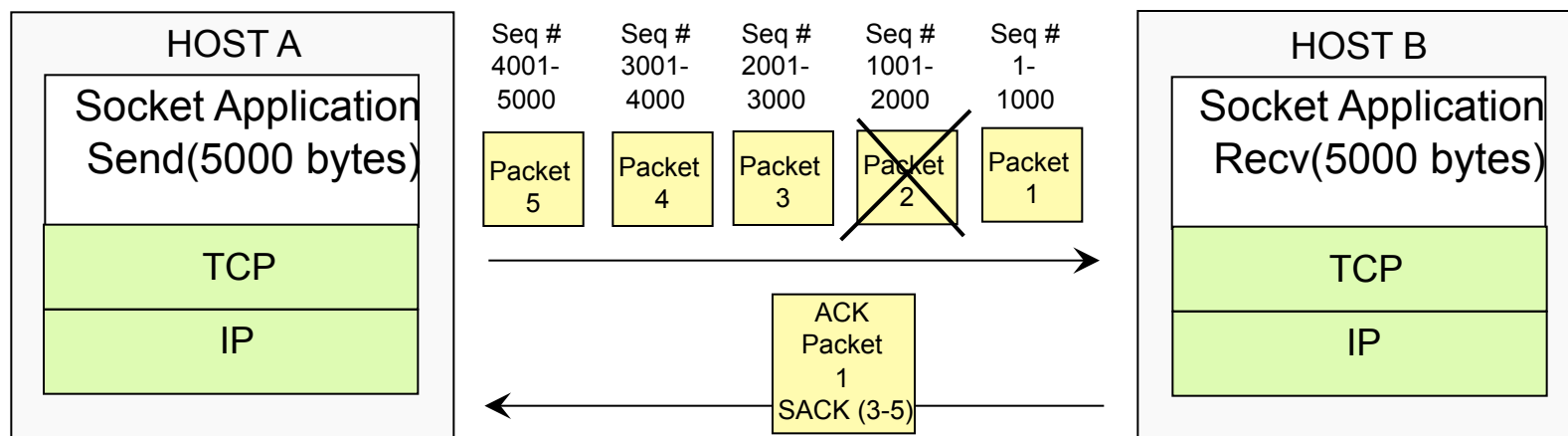
Note: Based on IBM internal benchmarks of modeled workloads in a controlled environment. The actual performance gains that users will experience may vary based on the user workload and configuration.

Background: TCP selective acknowledgement and retransmission



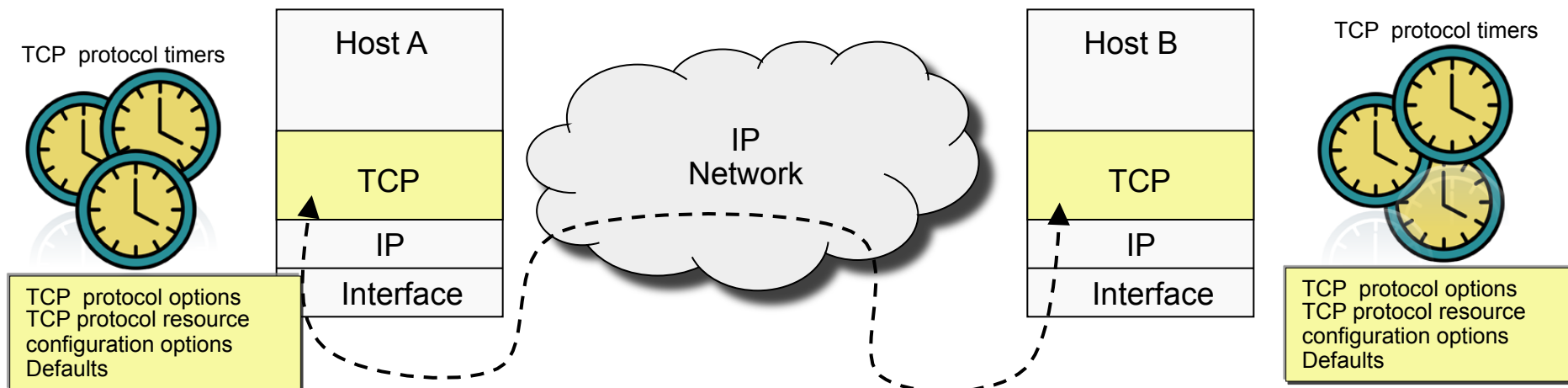
- TCP ack/retransmit processing
 - Receiving host detects gap in data
 - ACKs the last packet that was received in order for every packet received (multiple duplicate ACKs sent)
 - Sending host retransmits all packets after the acknowledged packet
- Results in unnecessary overhead
 - Sender retransmits more packets than necessary
 - Receiver burns cycles processing packets it has already received

V2R1 TCP selective acknowledgements and retransmission processing



- TCP ack/retransmit processing
 - Receiving host detects gap in data and ACKs the last packet that was received in order but includes information about other packets in the sequence that have been received (out of order segments)
 - Sending host can use the selective ACK information to only retransmit the missing packets (i.e. the gaps)
- Based on RFC 2018 (Selective TCP Selective Acknowledgment Options) and RFC 3517 (A Conservative Selective Acknowledgment (SACK)-based Loss Recovery Algorithm for TCP)
- SACK capability dynamically negotiated by TCP peers during connection establishment
- More efficient retransmission processing in networks with packet loss
 - But some additional CPU overhead under certain conditions
 - Environments where there is very little “real” packet loss
 - **AND INBPERF DYNAMIC NOWORKLOADQ** specified for OSA interfaces
 - SACK processing may take effect even without packet loss due to multi-processing of inbound traffic (TCP segments processed out of order)
 - Specifying INBPERF DYNAMIC WORKLOADQ solves this issue (eliminates race conditions for inbound TCP segments associated with bulk data connections)
- New configuration statements to enable/disable SACK processing
 - TCPCONFIG SELECTIVEACK|NOSELECTIVEACK option (Default: NOSELECTIVEACK)

Enhanced TCP protocol configuration options and default settings



- Background: The TCP protocol has several timers and resource utilization options
 - Influence the protocol's behavior in terms of resources consumed (e.g. memory for buffers, queued connections, etc.)
 - Timers that influence how long the protocol waits before taking an action (timing out connections, retransmitting packets, etc.)
- Problem:
 - Some of these timers and resource utilization options cannot be modified by users
 - Default settings for some of these options are not optimal for today's networks and workloads
- Solution:
 - Externalize key timers and controls via new configuration options in the TCP/IP profile
 - New TCPCONFIG options to control TIMEWAIT state intervals, Connection establishment timeouts, retransmission settings, KeepAlive settings, Nagle's algorithm, etc. (see Appendix for Details)
 - Modify default settings where possible to match best practices recommendations

Modified defaults for existing TCPCONFIG parameters in TCPIP profile

<i>TCPCONFIG parameter</i>	<i>Function</i>	<i>New Default</i>	<i>Old Default</i>
<i>SOMAXCONN</i>	<p>Controls how big the TCP backlog queue can be for a listening socket. The backlog value is specified on the listen() socket API.</p> <p>SOMAXCONN controls the maximum value that can be specified. In today's environments and workloads the previous default of 10 was too restrictive and needed to be modified by most customers.</p> <p>The default used to be 10 and is now changed to be 1024. Note: Memory is not pre-allocated based on this setting</p>	1024 connections	10 connections
<i>TCPRCVBUFRSIZE and TCPSENDBUFRSIZE</i>	<p>Control how much data can be stored by local TCP stack on a connection basis (data waiting to be read by application or sent to partner application)</p> <p>Defaults used to be 16K, now changed to 64K (note: memory is not allocated unless needed)</p> <p>These buffer sizes can also be programmatically changed via setsockopt() API on a socket basis.</p>	64K	16K

QDIO acceleration coexistence with IP filtering

- QDIOACCELERATOR

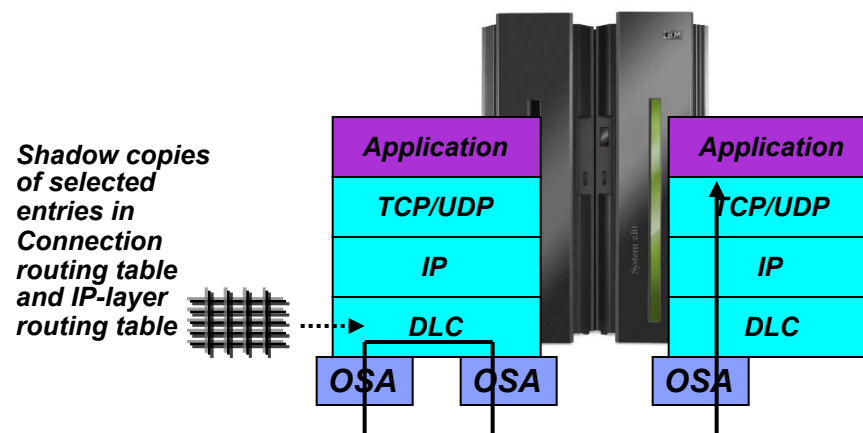
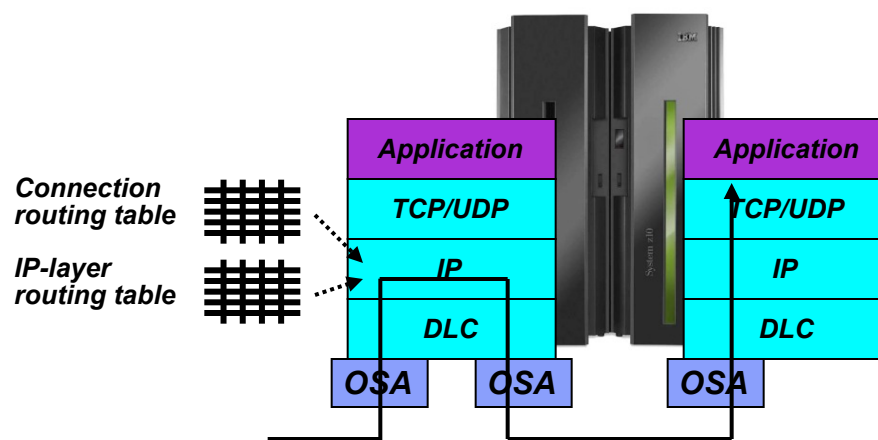
- Indicates that inbound packets that are to be forwarded should be routed directly between a HiperSockets device and an OSA-Express device in QDIO mode.
- Function is called “QDIO Accelerator” or “Hipersockets Accelerator”
- Affected packets are routed without touching the TCP/IP stack
- Improves performance and reduces processor usage for such workloads

- IPSECURITY

- Enables IP filtering in the TCP/IP stack

- It has not been possible to specify these two statements together

- The stack’s IP filters cannot be applied to QDIOACCELERATOR traffic since the routing occurs “below” the stack
- Specifying both in the same profile results in an error message



QDIO acceleration coexistence with IP filtering ...

- However, there are valid cases where it makes sense to specify QDIOACCELERATOR with IPSECURITY - cases where IP filtering is not needed for the QDIO Accelerator traffic:
 - The routed traffic is destined for a target that's doing its own endpoint filtering
 - IPSECURITY is only specified to enable IPSec on the local node
- V2R1 allows QDIOACCELERATOR to be specified with IPSECURITY in the TCPIP profile under certain conditions:

IP filter rules & defensive filter rules permit all routed traffic?	IP filter rules & defensive filter rules require routed traffic to be logged?	QDIO acceleration permitted?
N	N	N
N	Y	N
Y	Y	N
Y	N	Y
Sysplex Distributor traffic always forwarded		

z/OS V2R1 Communications Server Technical Update

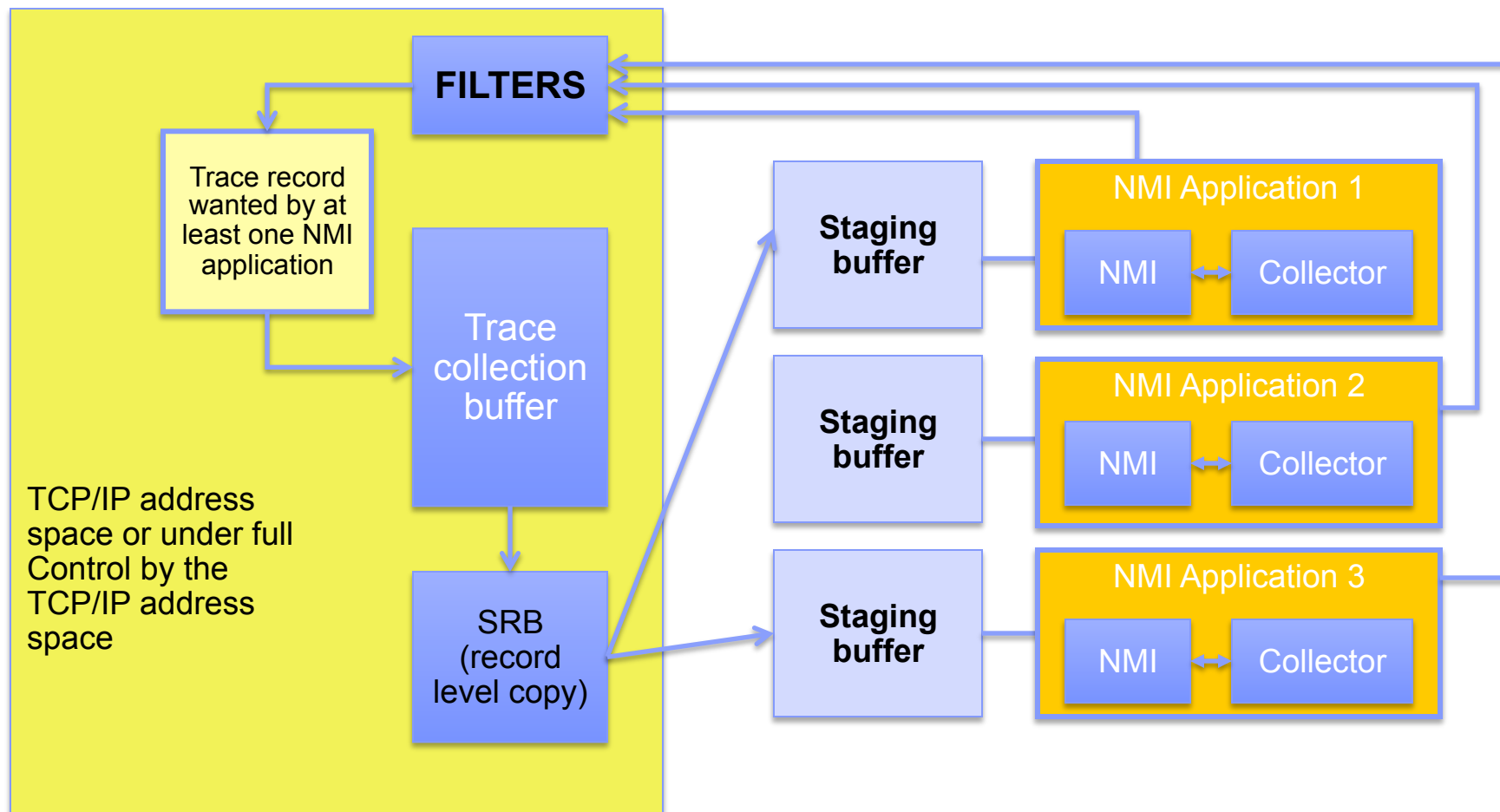
Application / Middleware / Workload enablement



Real-time application-controlled TCP/IP trace NMI

- Current TCP/IP network management interface (NMI) for real-time trace data
 - Allows multiple network management applications to obtain a copy of a packet trace that has been activated via console commands
 - Every application gets the same copy of the trace data
 - No ability to customize what data is collected by each application, a single trace may be active at any time
 - No ability to coordinate multiple distinct trace options
- New NMI that provides support for multiple concurrent and independent trace operations.
 - Applications use granular API-based trace open, filter definition, activation, deactivation, and trace close functions to control the traces they collect
 - Supports multiple independent trace operations running concurrently, each with its own set of trace options
 - Includes packet trace and data trace
 - Operate independently of each other and of the operator controlled trace operations that are started and controlled via MVS console commands
- Allows the inclusion of the following real-time NMI traces in a single trace data stream
 - Packet trace
 - Optionally includes IPsec data in the clear under RACF authorization
 - Data trace
 - Optionally includes AT-TLS data in the clear under RACF authorization
 - New RACF resource profiles for access to trace data

Real-time application-controlled TCP/IP trace NMI ...



- Notes about this diagram are on the next chart

Real-time application-controlled TCP/IP trace NMI ...

NOTES

- All trace records are stored in a TCP/IP-controlled collection buffer in TCP/IP 64-bit common storage
- Trace records copied by TCP/IP into a staging buffer for each NMI application
 - Circular set of trace records
 - Uses a 64-bit shared memory object for each staging buffer (NMI application specifies the size of the memory object)
 - One shared object per open trace instance
 - Obtained in application's address space
 - Stack address space gets affinity to shared object
 - Trace records copied to trace instance staging buffers by an SRB running in the TCP/IP address space
 - If records lost (for example because trace wraps before application collects records), application will be notified
- Application collector function can obtain trace data at any point in time
 - Does not have to wait for buffer to become full
 - Invokes an NMI function to obtain the data

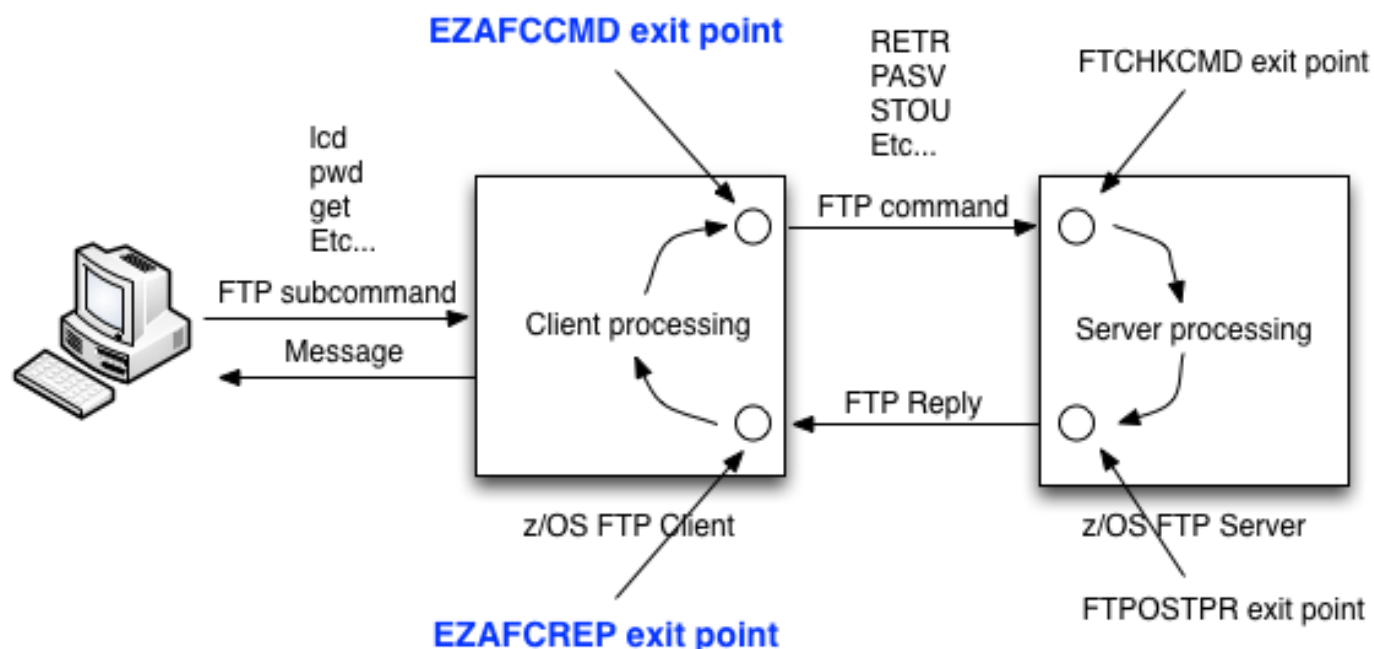
V2R1 FTP client security user exits

- You can use a variety of FTP server user exits to limit access to an FTP server
- However, system administrators currently have no way of controlling FTP client commands or other aspects of the processing done in the z/OS FTP client.
- Examples of FTP client controls that a system administrator might desire:
 - Preventing a dataset from being moved from the z/OS host (based on installation-specific criteria)
 - The ability to inspect or modify the dataset names specified by FTP client users for inbound and outbound file transfers
 - The ability to cancel an FTP client address space if that client is in the process of sending an “unauthorized” FTP command
- To address this, V2R1 implements two new FTP client user exit points as described on the next page
- These exits are defined and managed using z/OS dynamic exit services in order to ensure that only installation-approved exits are being used by FTP clients.

V2R1 FTP client security user exits ...

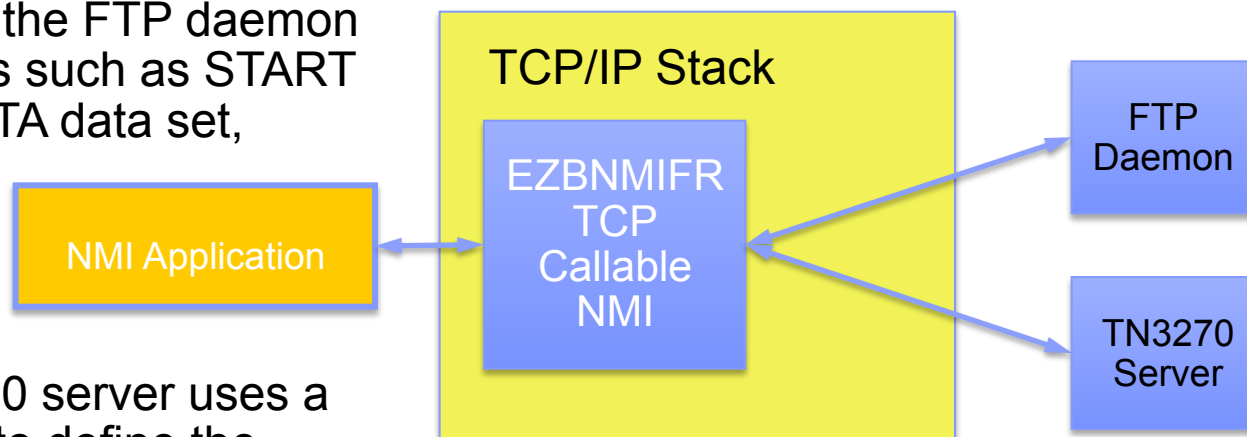
- EZAFCMD – FTP command user exit
 - called just before the FTP client sends an FTP command to the server
 - called before the command is converted to ASCII
 - Exit may inspect the command, modify the command arguments, reject the command or request the FTP client session be terminated

- EZAFCREP – FTP reply user exit
 - called whenever the FTP client receives a command reply over the control connection
 - called after the reply has been converted to EBCDIC
 - Exit may analyze the command results and request the FTP client session be terminated



NMI and SMF enhancements for TCP/IP applications

- Users can configure the FTP daemon with various methods such as START parameters, FTP.DATA data set, TCPIP.DATA data set, UNIX environment variables, etc.
- The z/OS CS TN3270 server uses a configuration profile to define the multiple listening ports for Telnet protocols, to define options, and for mapping client sessions to LU names for the interface with VTAM.
- z/OS V2R1 Communications Server provides a way for network management applications to obtain FTP daemon configuration data and TN3270 server configuration data using the existing TCP/IP callable NMI, EZBNMIFR, with new request types.
 - You can also obtain type 119 SMF records for FTP daemon configuration data and TN3270 profile information.



NMI and SMF enhancements for TCP/IP applications ...

NOTES

- The new NMI request type GetFTPDaemonConfig allows network management applications to programmatically obtain configuration data of active FTP daemons:
 - New TCP/IP callable NMI request type, GetFTPDaemonConfig - Get FTP Daemon Configuration Data
 - The NMI request requires a single filter which is the ASID of the FTP server whose configuration data is requested.
 - The NMI response includes a single record with all of the configuration data, which includes the [FTP.DATA](#) parameters, the START parameters, the UNIX environment parameters, and the TCPIP.DATA and [FTP.DATA](#) data set names.
 - Network management applications to call EZBNMCFG to get the configuration data of an FTP daemon can be written using C/C++ or Assembler.
- V2R1 also allows you to obtain a type 119 SMF record for FTP daemon configuration data.:
 - The new type 119, subtype 71 SMF record contains the FTP daemon configuration data and will be written after the FTP daemon finishes initialization and is listening on the listening port
 - Sections in this SMF record will be in same format as the corresponding sessions in the response buffer returned by the GetFTPDaemonConfig NMI request
 - A new server FTP.DATA parameter SMFDCFG is used to control whether to generate the above SMF record
 - As with other TCP/IP SMF records, this SMF record can either be written to an SMF data set, or written to a dataspace and accessed through the Network Management Real Time Interface for SMF Events (SYSTCPSM).

NMI and SMF enhancements for TCP/IP applications ...

NOTES

- The new NMI request type GetTnProfile allows network management applications to programmatically obtain complete TN3270 initial profile information.
- V2R1 also includes a new type 119, subtype 24 SMF record which provides the initial TN3270 profile information, as well as information about replacement of the profile caused by VARY TCPIP, Telnet, OBEYFILE processing.
 - As with other TCP/IP SMF records, this SMF record can either be written to an SMF data set, or written to a dataspace and accessed through the Network Management Real Time Interface for SMF Events (SYSTCPSM).
- Network management applications can use a combination of the GetTnProfile request and the new SMF 119 event records that are created during VARY TCPIP, Telnet, OBEYFILE command processing to monitor replacements of the Telnet profile.

Policy-based outbound routing of traffic that originates on z/OS

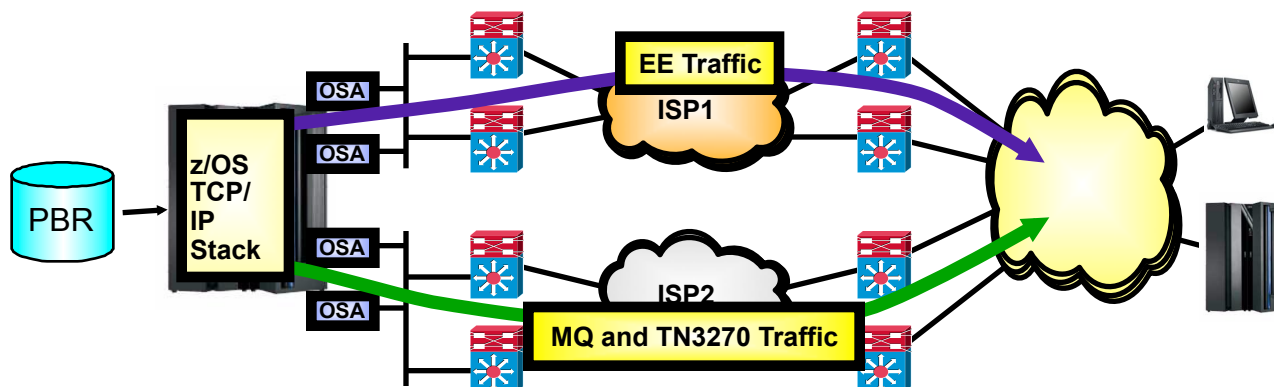
- **What does Policy Based Routing (PBR) do?**

- Choose first hop router, outbound network interface (including VLAN), and MTU
- Choice can be based on more than the usual destination IP address/subnet
 - With PBR, the choice can be based on source/destination IP addresses, source/destination ports, TCP/UDP, etc.

- **Allows an installation to separate outbound traffic for specific applications to specific network interfaces and first-hop routers:**

- Security related
- Choice of network provider
- Isolation of certain applications
 - EE traffic over one interface
 - TN3270 traffic over another interface
- PBR policies will identify one or more routes to use
 - If none of the routes are available, options to use any available route or to discard the traffic will be provided

*Enabled for IPv6 support
in V2R1*



PBR technologies are a great companion to VLAN technologies for separation of traffic over different networks or network providers.

z/OS V2R1 Communications Server Technical Update

Statements of Direction



Statement of Direction: End of support for TCP/IP legacy device drivers

z/OS V2.1 is planned to be the last z/OS release to provide software support for several TCP/IP device drivers. IBM recommends that customers using any of these devices migrate to more recent device types, such as OSA Express QDIO and Hipersockets. The TCP/IP device drivers planned to be removed are: Asynchronous Transfer Mode (ATM), Common Link Access To Workstation (CLAW), HYPERChannel, Channel Data Link Control (CDLC), SNALINK (both LU0 and LU6.2), and X.25.

Note: Support for SNA device drivers is not affected.

- A migration health check is planned that will determine if any of these legacy device types are defined in the TCP/IP Profile. This health check will be provided for V1R13 and V2R1 via APARs PI12977/OA44669 (V1R13) and PI12981/OA44671 (V2R1). These APARs are planned for availability in the second quarter of 2014.

Statement of Direction: z/OS Communications Server Internet mail applications: Sendmail and SMTPD

IBM intends to remove the Simple Mail Transport Protocol Network Job Entry (SMTPD NJE) Mail Gateway and Sendmail mail transports from z/OS Communications Server in the future. If you use the SMTPD NJE Gateway to send mail, IBM recommends you use the existing CSSMTP SMTP NJE Mail Gateway instead. CSSMTP provides significant functional and performance improvements.

The Sendmail client program can also be used to send mail messages; IBM plans to provide a replacement function using CSSSMTP as the SMTP transport, which will be designed so that it does not require application programming changes.

No replacement function is planned in z/OS Communications Server to support using SMTPD or Sendmail as a (SMTP) server for receiving mail for delivery to local TSO/E or z/OS UNIX System Services user mailboxes, or for forwarding mail to other destinations.

- See Appendix B for additional background information.

Please fill out your session evaluation

- z/OS V2R1 Communications Server Technical Update, Part 1
- Session # 14677
- QR Code:

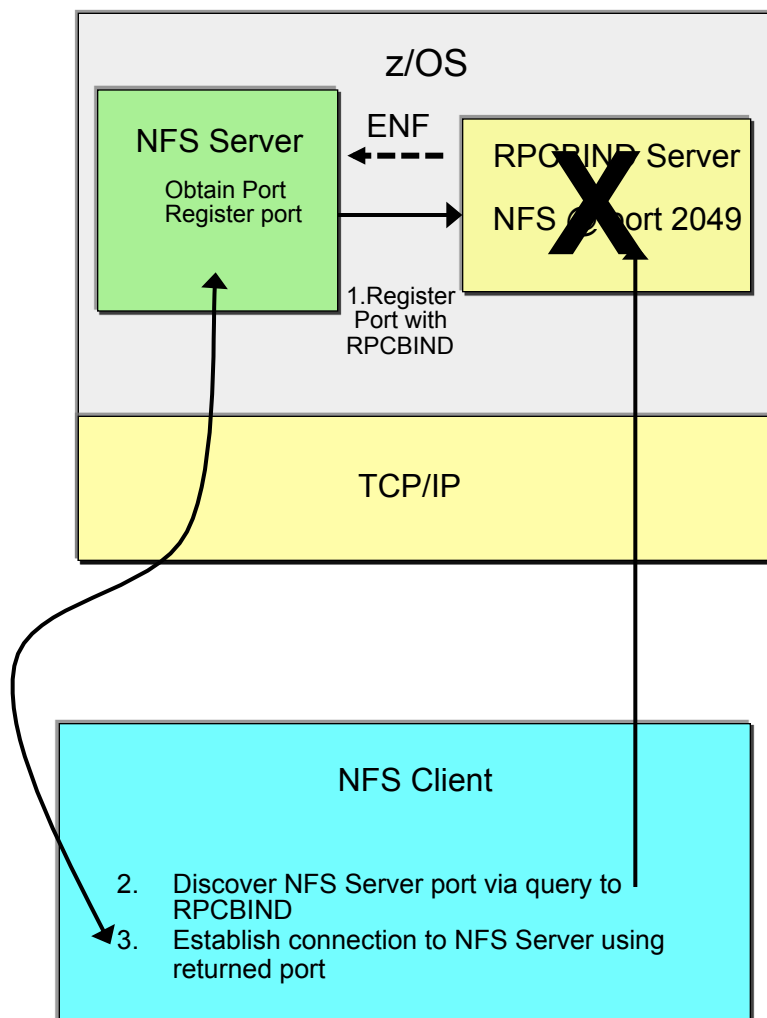


z/OS V2R1 Communications Server Technical Update

Availability



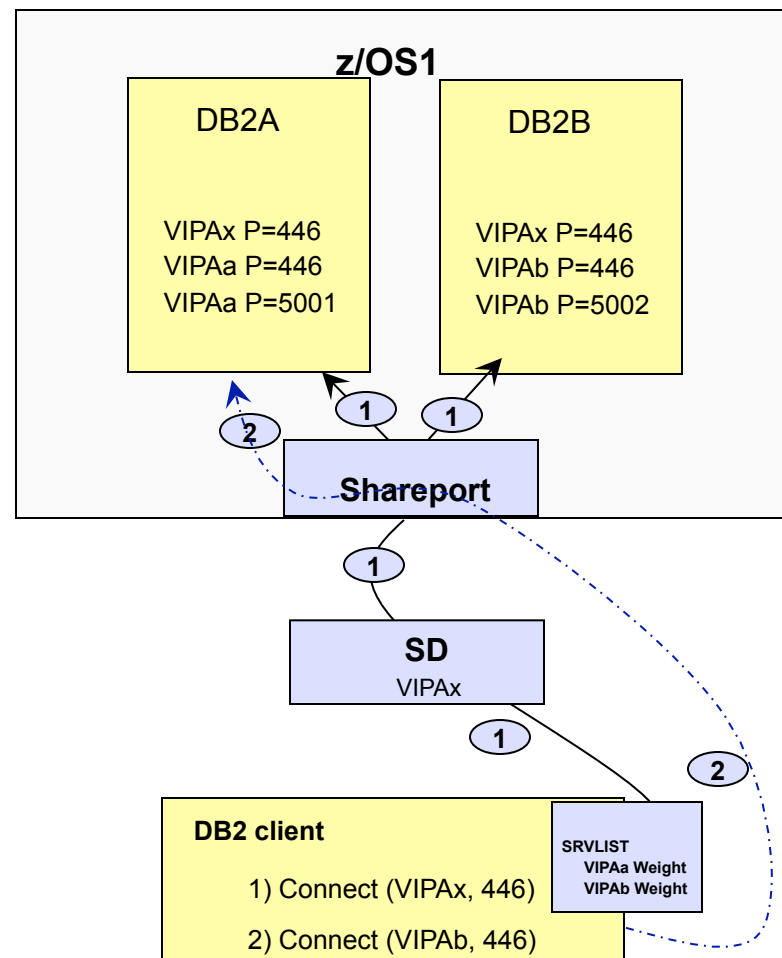
V2R1 RPCBIND recycle notification



- RPCBIND maintains a registry of RPC services and ports.
 - RPC servers register info at startup
 - RPC clients query RPCBIND to locate RPC servers
 - NFS client/server use RPCBIND
- When the RPCBIND server is recycled, all registration information is lost, preventing RPC clients from locating RPC services
 - Each RPC service must re-register
 - In the case of NFS this means recycling the NFS server which causes all existing NFS client connections to be terminated
- In V2R1, the RPCBIND server now raises an ENF signal when either RPCBIND is started or is stopping allowing RPC servers to dynamically re-register all their RPC ports with RPCBIND
- This enhancement only applies to RPCBIND – NOT PORTMAP
 - PORTMAP is a an IPv4-only subset of RPCBIND
 - Consider moving to RPCBIND if you still use PORTMAP

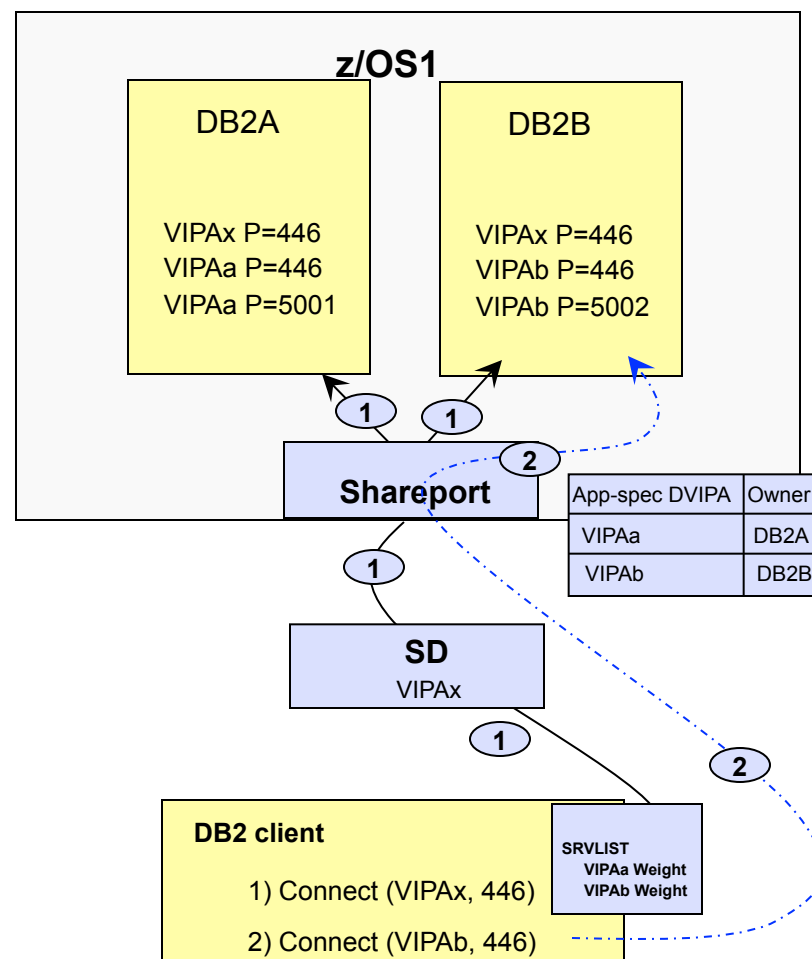
Dynamic VIPA affinity support – Problem Scenario

- Multiple DB2 members deployed in the same z/OS image
- Currently 2 methods for creating Dynamic VIPAs to represent each member (Member-specific DVIPAs)
 1. BIND specific approach: Using the TCP/IP profile PORT BIND specific support each member binds its listening socket to a specific member DVIPA
 - *All is well*, incoming connections get routed to the appropriate member
 2. Dynamic VIPAs configured via DB2 BSDS (Bootstrap dataset)
 - DB2 programmatically creates the DVIPAs during initialization and *binds listening sockets to INADDR_ANY*
 - o Allows each DB2 member to be reached using any IP address (including the IPv4 and IPv6 DVIPAs, DB2 Location Alias IP addresses)
 - o However if multiple DB2 members in the same z/OS system use the BSDS method, *incorrect routing of incoming connections is possible*
 - o Resulting in suboptimal load balancing



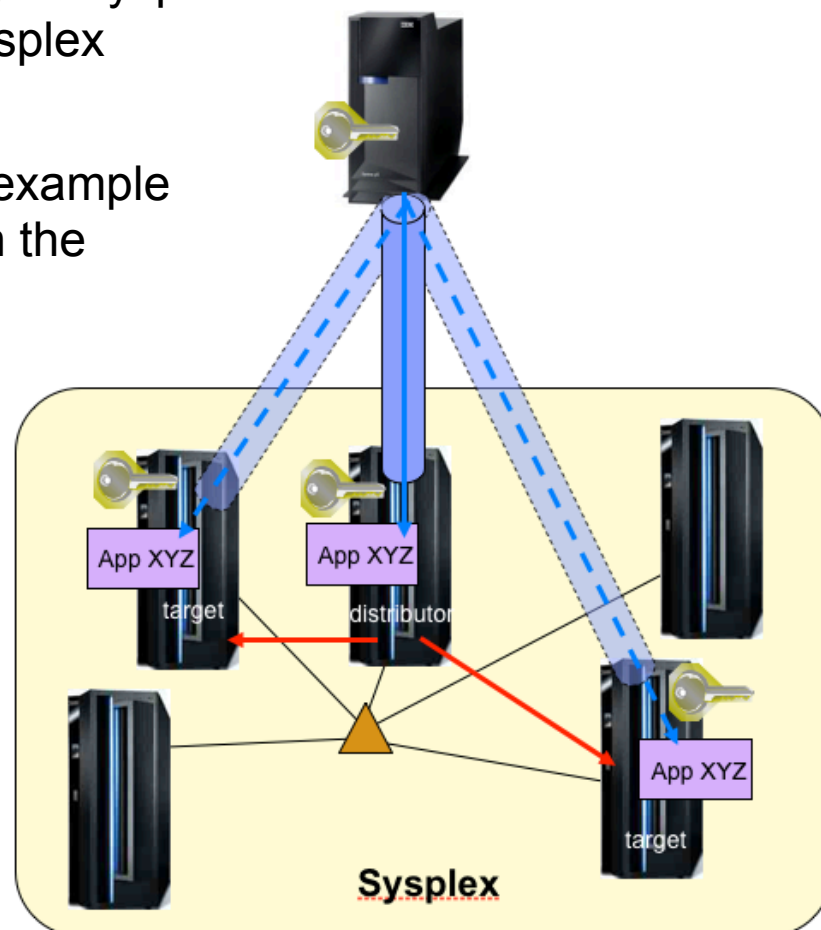
Affinity for application-instance DVIPAs - Solution

- **Provide a capability to create an application instance DVIPA with affinity**
 - DVIPA affinity determined by creating address space (i.e. DDF instance)
 - Incoming connections to an “affinity” DVIPA are handled in a special manner by SHAREPORT processing
 - When multiple listening sockets for the target port are available find the listening socket owned by the address space that created the DVIPA
 - If an address space with affinity is not found with a listening socket on the target port then route the connection to any listening socket that can accept it (i.e. bound to INADDR_ANY)
 - Allows the DVIPA to be used by non-DB2 applications, such as incoming FTP connections to the member-specific DVIPA address
 - Only works while the DVIPA is still active
- **Support to create DVIPA with affinity is provided on**
 - Socket APIs (SIOCSVIPA and SIOCSVIPA6 IOCTLS)
 - MODDVIPA utility program
- **Allows a sysplexWLB connection using a BSDS DVIPA (INADDR_ANY approach) to land on the intended DB2 member even when multiple DB2 members coexist on the same z/OS image.**
 - Will require new DB2 exploitation support
 - APAR PI08208 (available on DB2 V10 and DB2 V11)



Sysplex-Wide Security Associations for IPv6

- Sysplex-Wide Security Associations (SWSA) allow IPsec-protected traffic to be distributed through a sysplex while maintaining end-to-end security to all sysplex endpoints.
- Security associations and characteristics (for example encryption) are moved and distributed through the sysplex with the target programs
 - VIPA Takeover: Ability of an SA to follow a DVIPA when it moves from one stack to another
 - Distribution: Ability of a target stack to use an IPsec SA that was negotiated on its behalf by the Sysplex distributor stack.
- Previously supported only for IPv4
- V2R1 completes the SWSA capability by adding IPv6 support

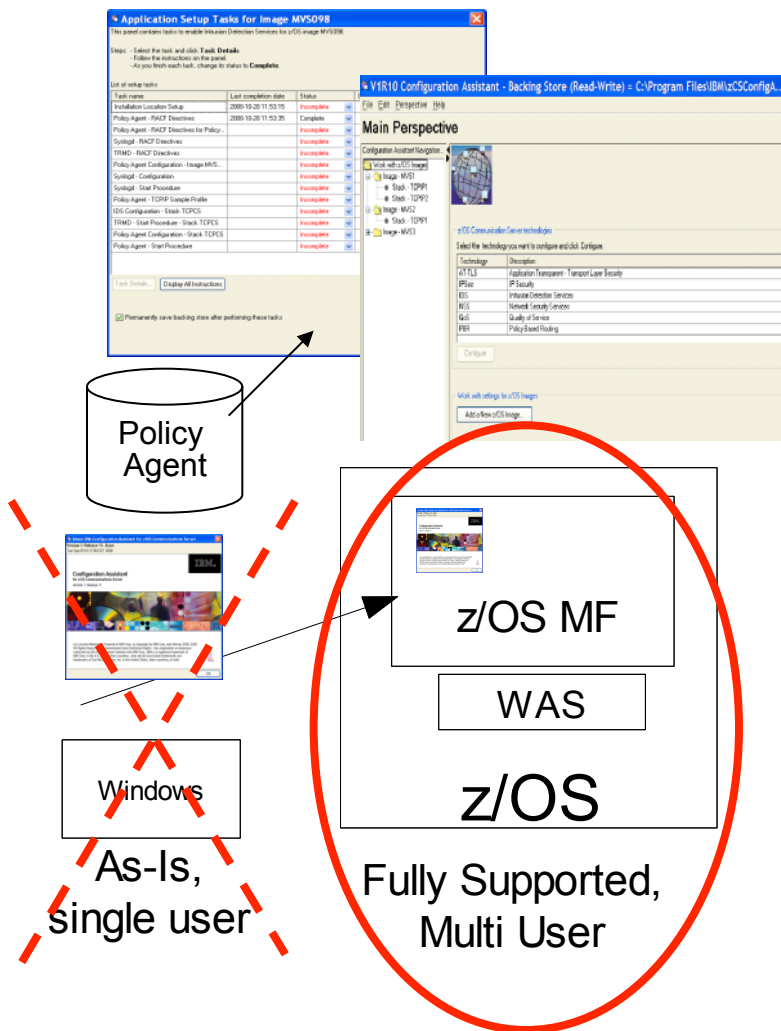


z/OS V2R1 Communications Server Technical Update

Simplification



Review: IBM Configuration Assistant for z/OS Communications Server

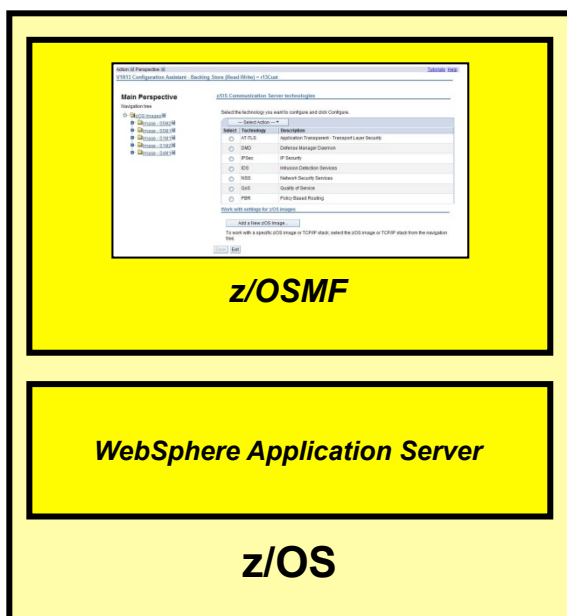


- As of z/OS V1R11, IBM Configuration Assistant for z/OS Communications Server is integrated with z/OS Management Facility (z/OSMF)
 - z/OSMF version is integrated into the product and runs on z/OS.
 - z/OSMF version is officially supported.
 - Supports policy definitions for IPSECURITY (IP filters, IPsec), AT-TLS, IDS, Network QoS, etc.
- The standalone Windows version is still available, but is made available as-is, without any official support:
 - <http://tinyurl.com/cgoqsa>

Statement of Direction: IBM Configuration Assistant for z/OS Communications Server

July, 2011: z/OS V1.13 is planned to be the final release for which the IBM Configuration Assistant for z/OS Communications Server tool that runs on Microsoft Windows will be provided by IBM. This tool is currently available as an as-is, nonwarranted web download. Customers who currently use Windows-based IBM Configuration Assistant for z/OS Communications Server tool should migrate to the z/OS Management Facility (z/OSMF) Configuration Assistant application. The IBM Configuration Assistant for z/OS Communications Server that runs within z/OSMF is part of a supported IBM product and contains all functions supported with the Windows tool.

V2R1 Configuration Assistant for z/OS Communications Server



***Interface for
Communication Server
policy-based
definition, installation
and activation***

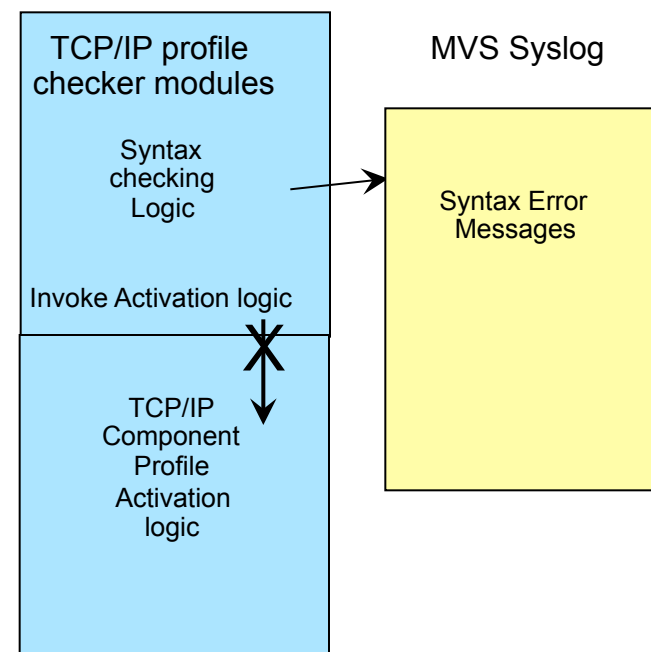
- Complete rewrite from AUIML (Abstract User Interface Markup Language) to Javascript using Dojo toolkit
- Consistent look and feel with z/OSMF
- Significantly better performance
 - More work done on the client's browser = much faster response time
 - Less processing on the z/OSMF host = much lower Config Assistant CPU utilization
- Reduced number of clicks for some common tasks
- Uses Web 2.0 model based on RESTful services defined between the client side browser and the Config Assistant running on the z/OSMF server
- Works with your existing backing store

Check TCP/IP profile syntax without applying configuration changes

- Prior to V2R1, there has been no easy way to validate TCP/IP profile syntax
 - Original TCP/IP profile during stack activation
 - V TCPIP,,OBEY processing
- Syntax errors can lead to undesirable results:
 - Partial profile put into effect - could lead to unintended outage
 - Changes caused by earlier statements might need to be undone before processing a repaired profile
- New command in V2R1 triggers the reuse of the existing TCP/IP profile parser to check the syntax:

V TCPIP,,SYNTAXCHECK,dsname

```
EZZ0065I VARY SYNTAXCHECK COMMAND BEGINNING
...
<error messages as syntax errors encountered>
...
EZZ0064I VARY SYNTAXCHECK FOUND ERRORS: SEE PREVIOUS MESSAGE
EZZ0065I VARY SYNTAXCHECK COMMAND COMPLETE
```



Check TCP/IP profile syntax without applying configuration changes ...

- The VARY SYNTAXCHECK command requires an active TCP/IP stack at the same level as the intended target system
 - Can be a development or test system, does not need to be issued on the target system
 - Note: If system symbolics are being used they will be resolved based on the system symbolic configuration of the system the command is issued on
- Can limit access to the command via SAF profiles
- Processes all INCLUDE files specified (even nested ones, just like OBEYFILE)
- Will only flag syntax errors, not semantic (configuration) errors
 - Does not validate TCP/IP profile statements against the current active configuration (e.g., issuing a DELETE PORT for a port that is not currently defined will not be flagged)
- No updates are applied to the active configuration



IPv4 INTERFACE statement for HiperSockets and static VIPAs



- In z/OS V1R4, a new TCP/IP PROFILE statement to define IPv6 network interfaces was introduced – the INTERFACE statement
 - One statement for all interface-related attributes for IPv6 network interfaces
 - IPv4 interfaces continued to require use of DEVICE, LINK, HOME, and optionally BSDROUTINGPARMS statements to define all the attributes of an IPv4 network interface
- z/OS V1R10 extended the use of the INTERFACE statement to IPv4 QDIO network interfaces
 - Non-QDIO IPv4 network interfaces continued to require the old configuration syntax
- z/OS V2R1 allows the INTERFACE statement to be used to configure IPv4 HiperSockets and static VIPAs
 - Provides a more straightforward way of configuring the source VIPA for these IPv4 interfaces.
 - Allows for configuration of multiple VLANs from the same TCP/IP stack for a single HiperSockets CHPID for both IPv4 and IPv6.
 - Consider migrating to INTERFACE statements where supported
 - DEVICE/LINK statements should only be required for legacy DLC's (LCS, Claw, etc.)

```
INTERFACE HSINTF1
  DEFINE   IPAQENET
  IPADDR   200.16.1.1/24
  CHPID    FE
  SOURCEVIPAINTERFACE VIPA4811L
```

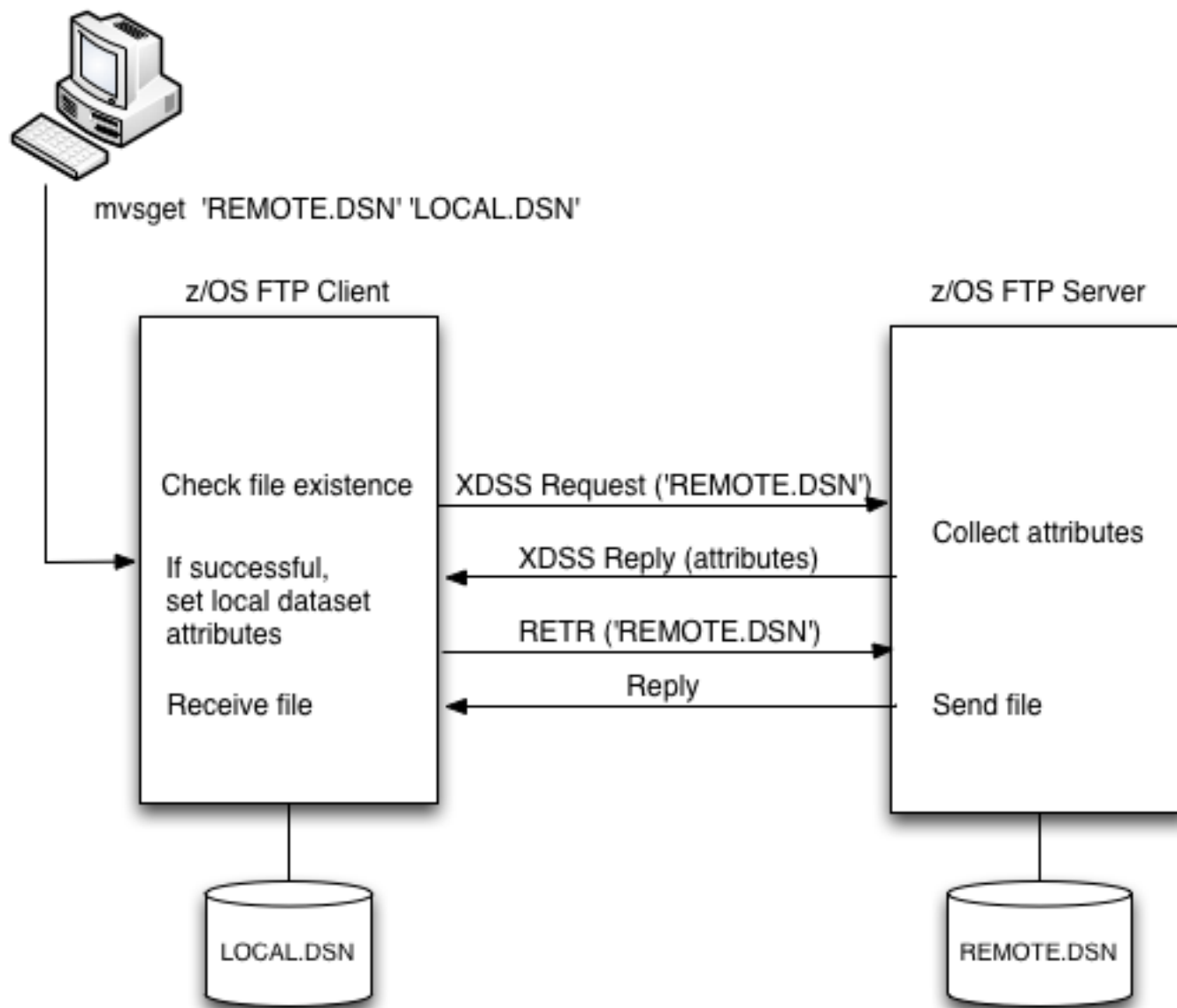

Simplify FTP transfer of datasets between z/OS systems

- When you log into the z/OS FTP server with the z/OS FTP client to get an MVS data set from the server, you have to know a lot about the source data set. You also have to follow many steps to complete the transfer
 - Determine attributes of the source data set:
 - Record format information (LRECL, RECFM, BLKSIZE)
 - SPACE information (allocation units, primary and secondary allocation)
 - Data set type (sequential, PDS or library – and if PDS: number of directory blocks)
 - Issue one or more LOCSITE subcommands to set these attributes on the client side
 - For sequential data sets: issue a GET subcommand
 - For PDS or library data sets: issue an Imkdir subcommand, followed by an MGET * subcommand
 - The transfer might result in the source and target data sets having mismatched attributes if the target data set already exists. This increases the likelihood of data loss due to wrapping, truncation, space constraints, and so on
- The same problem exists when you want to transfer MVS data sets from the z/OS FTP client to the z/OS FTP server.

Simplify FTP transfer of datasets between z/OS systems ...

- V2R1 simplifies FTP data transfers between z/OS systems through three new commands that hide the complex interactions under the covers
 - XDSS command
 - used internally by FTP to get the attributes of the remote MVS data set and send them back to the z/OS FTP client in a 200 reply
 - MVSGet and MVSPut subcommands
 - Encapsulate the complex interactions for transferring an MVS data set between z/OS systems
 - FTP automatically extracts the data set attributes of the source data set and then apply them to the target system before allocating the target data set
 - Allows you to reallocate the existing target data set instead of replacing it. This enhances the reliability of data set transfers by ensuring that the source and target data set attributes match
 - Lets you transfer a PDS or library as a whole – something you couldn't do before V2R1.

Simplify FTP transfer of datasets between z/OS systems ...



z/OS V2R1 Communications Server Technical Update

Security



z/OS Application Transparent TLS overview

▪ Stack-based TLS

- TLS process performed in TCP layer (via System SSL) without requiring any application change (transparent)
- AT-TLS policy specifies which TCP traffic is to be TLS protected based on a variety of criteria
 - Local address, port
 - Remote address, port
 - Connection direction
 - z/OS userid, jobname
 - Time, day, week, month

▪ Application transparency

- Can be fully transparent to application
- An optional API allows applications to inspect or control certain aspects of AT-TLS processing – “application-aware” and “application-controlled” AT-TLS, respectively

▪ Available to TCP applications

- Includes CICS Sockets
- Supports all programming languages except PASCAL

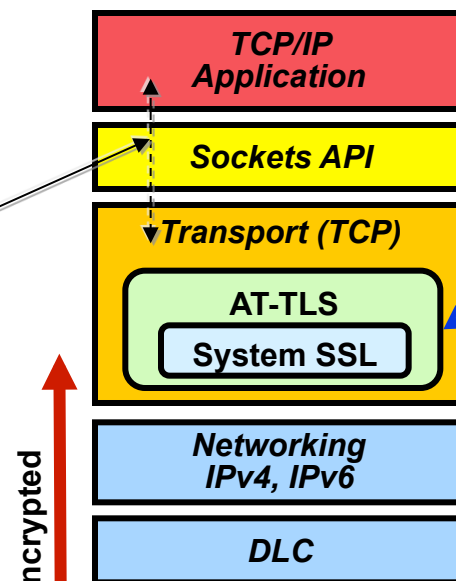
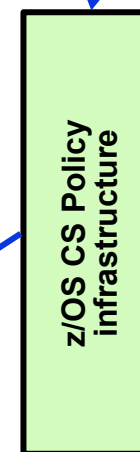
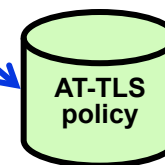
▪ Supports standard configurations

- z/OS as a client or as a server
- Server authentication (server identifies self to client)
- Client authentication (both ends identify selves to other)

▪ Uses System SSL for TLS protocol processing

- Remote endpoint sees an RFC-compliant implementation
- interoperates with other compliant implementations

AT-TLS policy administrator using Configuration Assistant



Key AT-TLS exploiters:

- FTP, TN3270, DB2, IMS Connect, IMS SOAP GW, CICS Sockets, NJE IP, RACF (RRSF over TCP/IP), CIMOM, Netview

AT-TLS support for TLS v1.2 and related features

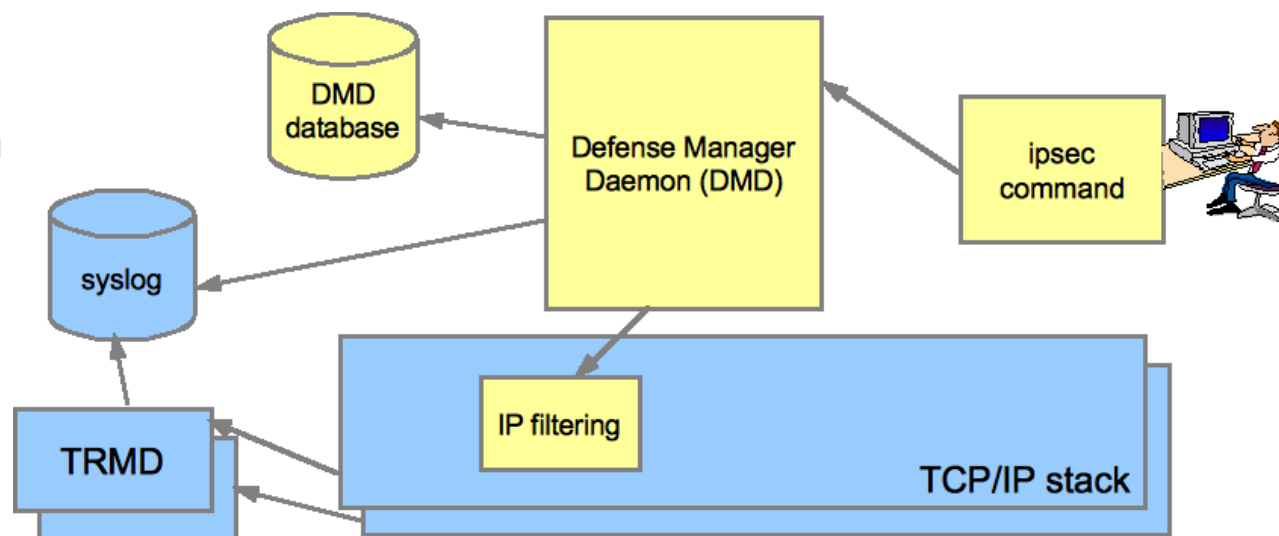
- Transport Layer Security (TLS) Renegotiation Extension (RFC 5746):
 - Provides a mechanism to protect peers that permit re-handshakes
 - When supported, it enables both peers to validate that the re-handshake is truly a continuation of the previous handshake
- Support Elliptic Curve Cryptography (ECC)
 - Twenty new ECC cipher suites
 - ECC cipher suites for TLS (RFC 4492)
- TLS Protocol Version 1.2 (RFC 5246):
 - Twenty-one new cipher suites
 - 11 new HMAC-SHA256 cipher suites
 - 10 new AES-GCM cipher suites
 - Requires new System SSL support
 - Addresses NIST SP800-131a requirements
- Support for Suite B cipher suites
 - TLS is required
 - ECC 128-bit or 192-bit cipher suites are required

- TLS v1.2 support is also available on V1R13 via APARs OA39422 and PM62905



Limit defensive filter logging

- Defensive filtering provides a mechanism to install temporary defensive filters into a TCP/IP stack to block a specific attack or pattern of attacks



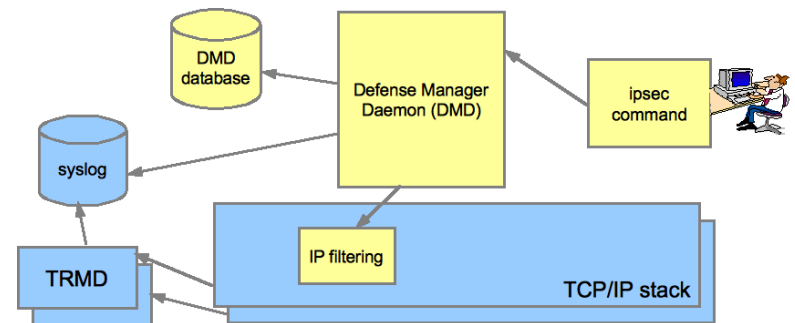
- Two modes of defensive filters:

- Blocking mode – denies packets that match the defensive filter
- Simulate mode – for a packet that matches the defensive filter, a log record is written to syslogd to indicate that the packet would have been denied if this were a blocking filter, but filtering of packet continues
 - Often used initially by customers to gauge the effect of implementing defensive filtering before enabling blocking mode
 - “EZD1722I Packet would have been denied by defensive filter” logged for each packet that matches this filter

Limit defensive filter logging...

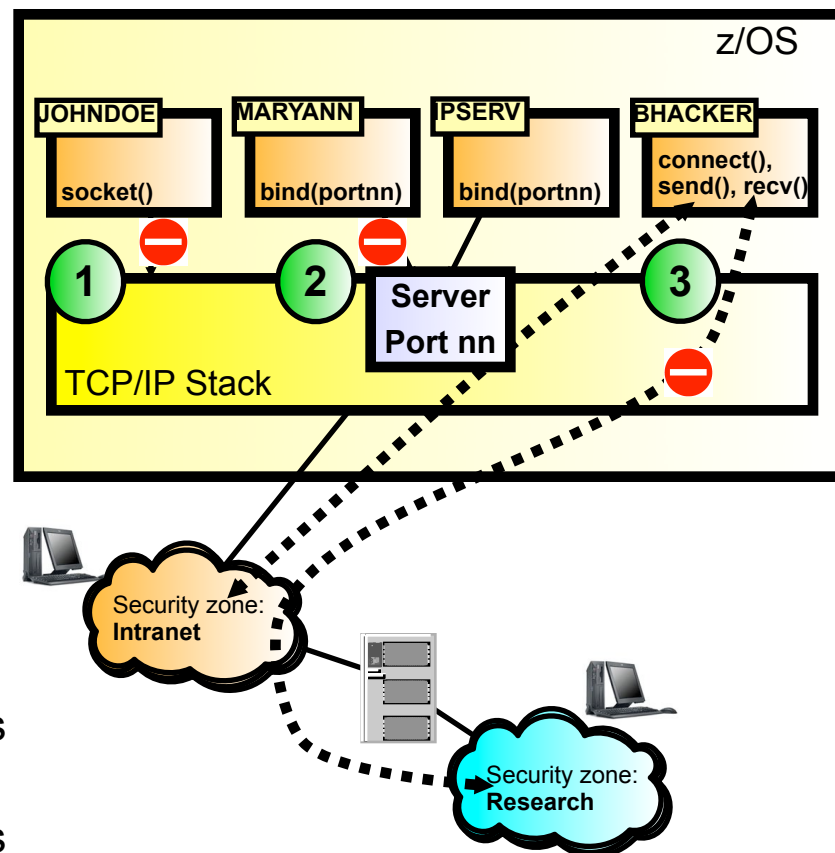
- Per-packet logging can flood syslogd during an attack
 - In blocking mode, logging can be turned on or off
 - In simulate mode, logging is always on, with a log record generated for each packet that matches the filter
 - No mechanism to limit syslogd output

- V2R1 adds a user-specified limit for the number of defensive filter messages written to syslogd over a 5-minute interval
 - Limit is per defensive filter (not across all defensive filters)
 - Limits number of EZD1721I and EZD1722I defensive filter messages
 - Value from 0 – 9999
 - 0 indicates “no limit” - message written to syslogd for each packet that matches the defensive filter
 - 1 – 9999 – indicates limit to be applied for defensive filter
 - Number of suppressed messages reported per defensive filter every five minutes (and when a defensive filter expires)



Improved auditing of NetAccess rules

- NetAccess provides the ability to control z/OS user access to certain security zones (networks, subnetworks, and hosts)
- Via NETACCESS statement in TCP/IP
- Access to security zone allowed if user permitted to SAF resource (SERVAUTH class: NETACCESS)
- Results from SAF calls to check if a user can access a security zone are cached
 - Only first access check for a user to a security zone results in a SAF call
 - No SAF call for subsequent access checks for user for same IP address
 - No SAF call for subsequent access checks for user to different IP addresses in same zone
- NetAccess provides a log string on all calls made to SAF to check a user's access to a SAF resource profile, and SAF includes that log string in its audit records (for example, RACF SMF record)



Improved auditing of NetAccess rules ...

- Customers have asked for more control of caching to provide more audit details
 - SAF audit records are only written when a SAF call is made
 - Security auditors need audit records that are inhibited by caching
- In V2R1, a new parameter is added to NetAccess to control caching:

Parameter	Description	Effect on SAF SMF records	Effect on auditing behavior
CACHEALL	<ul style="list-style-type: none"> ▪ Results from all NETACCESS SAF calls are cached, regardless of whether the user is permitted access to the zone ▪ This is the default and is the same as previous caching behavior 	Audit record written for only the first access check made for a user to each security zone	Allows for auditing of only the first access check made for each user to each security zone
CACHEPERMIT	Results from NETACCESS SAF checks are cached when the user is permitted access, but not when the user is denied access to the zone	<ul style="list-style-type: none"> ▪ Audit record written for only the first access check made for a user to each security zone to which user is permitted ▪ Audit record written for all access checks made for a user to each security zone to which user is denied 	<ul style="list-style-type: none"> ▪ Allows for auditing of only the first access check made to zones to which the user is permitted ▪ Allows for auditing of all access checks made to zones to which the user is denied
CACHESAME	<ul style="list-style-type: none"> ▪ Same as CACHEPERMIT ▪ However ▪ The cache is used by a socket only as long as the user and the IP address being accessed remain unchanged 	<ul style="list-style-type: none"> ▪ Same as CACHEPERMIT Plus ▪ Audit record written for next access check after socket user or remote IP address being used by the socket changes 	Allows for auditing of all access checks made to all zones (both permitted and denied) except for successive checks by a socket for the same user and the same IP address in a permitted security zone

Improved auditing of NetAccess rules ...

- SAF audit record contains user ID and resource profile name, but not the IP address that is being accessed
 - The security zone associated with the resource profile can contain multiple IP addresses
 - No record of which IP addresses within the security zones are being accessed
- Security auditors need the IP address information
 - Especially important when access is denied
- In V2R1, all SAF calls made for NetAccess include the IP address that triggered the call

Improved FIPS 140 diagnostics

- z/OS Communications Server components that offer a FIPS-140 operational mode:
 - IKED, NSSD, and AT-TLS (configured in FIPS-140 mode)
- In FIPS-140 mode, components must call z/OS cryptographic modules for all cryptographic operations
 - Must call ICSF and System SSL (configured in FIPS-140 mode)
 - Internal routines disabled
 - Direct hardware calls disabled
- Starting in V2R1, System SSL in FIPS-140 mode must call ICSF
 - Change made to satisfy FIPS-140 requirements, eliminate redundancy, and improve efficiency
 - Applies to FIPS-140 mode only
- As a result, components calling System SSL in FIPS-140 mode now require ICSF
 - IKED and NSSD will fail to initialize if ICSF is not active, and AT-TLS policy groups will be installed but inactive if ICSF is not active
 - In V2R1, z/OS CS has new messages to indicate ICSF status during IKED and NSSD initialization, and during the installation of AT-TLS policy groups

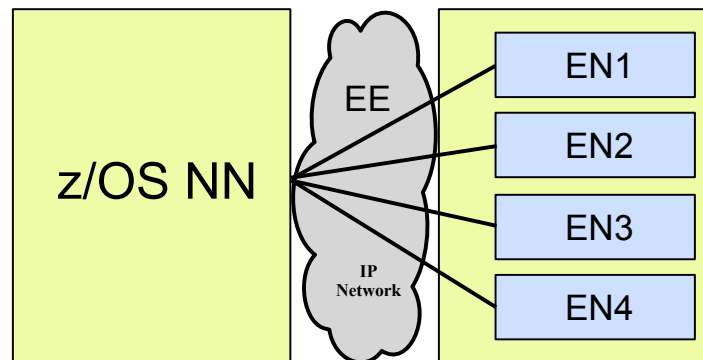
z/OS V2R1 Communications Server Technical Update

Enterprise Extender, SNA, and miscellaneous enhancements



EE/SNA

- V1R11 introduced Progressive Mode ARB flow control to address the tendency of HPR's Responsive Mode ARB to be very sensitive to minor variations in packet round-trip time or unpredictability in response time from the RTP partner node. If the partner suddenly becomes CPU constrained, even for a short period, throughput and response time can be degraded. For example:
 - Partner node has a shortage of CPU availability, memory, or network bandwidth
 - Partners in a virtual server environment on a single hardware platform cannot guarantee consistent response time
- Progressive mode ARB implemented several small changes to the flow control rules to improve responsiveness in a CPU-constrained environment
 - Both partners must agree to use progressive mode ARB
 - Limited to single-hop pipes over an EE connection (including two-virtual-hop connection network paths)
 - HPREEARB = PROGRESS can be specified on an EE PU in a switched major node, on a connection network GROUP in the EE XCA major node, or on the EE model PU in a model major node
- **V2R1 added the HPREEARB parameter to the GROUP statement for pre-defined EE connections**
 - Provides the ability to specify the HPREEARB parameter on the switched major node GROUP statement for pre-defined EE connections



EE/SNA ...

- IPv6 support for EE's IPADDR operand
 - Prior to V2R1, EE's IPADDR parameter is IPv4-only, with the assumption that IPv6 addresses will be provided via HOSTNAME resolution. V2R1 adds IPv6 support for the IPADDR start option, XCA, and switched PU parameters
- PSRETRY Enhancements
 - Provide a PSRETRY option to look for a new route after a local topological change. This is sort of the inverse of link-inop-driven path switch, and should provide better function and performance than the current timer-only mechanism.
- Display EE command enhancements
 - A new CPNAME filter is provided for DISPLAY EE to request all active EE connections to a given partner CP name
- Enhanced TRS traces
 - Provides additional internal traces to be used by z/OS CS service in diagnosing APPN routing problems
- The TSO DNET command has been removed from z/OS CS

```
VTAM Start Option List
...
IPADDR=2000::67:1:2
...
```

Miscellaneous items

- The Communications Server system resolver will now start even if errors are detected with statements in the resolver setup file.
 - The resolver will also start if the resolver setup file does not exist or cannot be accessed by the resolver.
 - This allows your TCP/IP stacks and other applications dependent on resolver processing to continue their initialization despite any resolver setup file errors.
- OMPROUTE change to turn on HELLO_HI by default
 - Enable processing of OSPF hello packets at a high priority by default. Adds a new global configuration parameter Enable_Hello_Hi with YES and NO values to the GLOBAL_OPTIONS statement in the OMPROUTE configuration file.
- Netstat socket creation time
 - The NETSTAT ALL output is updated to include a start date and time for a socket connection
- The EZZ4310I messages that report device failures are supplemented by additional messages that further describe the failure
- In V2R1, z/OS Communications Server provides a mechanism that allows an application to issue a synchronous or asynchronous receive socket API call that completes only when a TCP connection is terminated.

z/OS V2R1 Communications Server Technical Update

Appendix A: Miscellaneous additional enhancements



New TCPCONFIG parameters in TCPIP profile

New TCPCONFIG parameter	Function	Range	Default
FINWAIT2 (modified, not new)	Existing Parameter , controls how long (number of seconds) TCP connections stay in a FINWAIT2 state. Large number of connections in this state consume system memory. Change is in the range of values that can be specified.	1-3,600 seconds (previous range was 60-3,600 seconds)	600 seconds (unchanged)
TIMEWAITINTERVAL	Controls how long connections are kept in a TIMEWAIT state (occurs after a TCP connection is closed – close initiator goes into this state). Large numbers of connection in TIMEWAIT state consume system memory	0-120 seconds	60 seconds
MAXIMUMRETRANSMITTIME	TCP/IP <i>stack wide</i> control for maximum TCP retransmission interval. Effective if no other existing retransmission configuration options (i.e. MAXIMUMRETRANSMITTIME on BEGINROUTES/GATEWAY, ROUTETABLE or Max_Xmit_Time on OSPF_INTERFACE and RIP_INTERFACE)	0-999.99 seconds	120 seconds 15 attempts
RETRANSMITATTEMPTS	New control, specifies the total number of times a segment is retransmitted before aborting a TCP connection	0-15 attempts	15 attempts (same as today)

New TCPCONFIG parameters in TCPIP profile (cont)

New TCPCONFIG parameter	Function	Range	Default
CONNECTTIMEOUT	Specifies the total amount of time to allow before the initial connection times out (for outbound connections from z/OS).	5-190 seconds	75 seconds (previous internal setting was around 3 minutes)
CONNECTINITINTERVAL	Specifies the initial retransmission time out interval in milliseconds	100-3,000 ms	3,000 ms (3 seconds)
KEEPALIVEPROBES	Additional control related to TCP <i>Keepalive Interval</i> that can be specified on TCPCONFIG. This new control limits the number of keepalive probes that are sent after the keepalive timer has expired and the connection is not responsive. Note: Applications issuing setsockopt() TCP_KEEPALIVE will override this TCPCONFIG setting	1-10 probes	10 probes (same as previous releases)
KEEPALIVEPROBEINTERVAL	Related to KEEPALIVEPROBES, indicates the interval between probes. Note: Applications issuing setsockopt() TCP_KEEPALIVE will override this TCPCONFIG setting	1-75 seconds	75 seconds (same as previous releases)
TCPMAXSENDBUFRSIZE	The maximum send buffer size, which is between the value that is specified on the TCPSENDBUFRSIZE parameter and 2M. The default value is 256K.	Value on TCPSENDBUFRSIZE – 2MB	256K

New TCPCONFIG parameters in TCPIP profile (cont)

<i>New TCPCONFIG parameter</i>	<i>Function</i>	<i>Range</i>	<i>Default</i>
<i>NONAGLE and NAGLE</i>	<p>Nagle's algorithm prevents TCP from sending data packets which are smaller than a full segment unless all previously sent data has been ACKed. For most workloads this is not a problem. However, for some workloads that perform multiple small sends without receiving any inbound data this can lead to significant latency issues.</p> <p>The setsockopt() TCP_NODELAY option allows applications to disable Nagle's algorithm for a given connection.</p> <p>NONAGLE disables Nagle's algorithm for all TCP connections.</p> <p>NAGLE enables Nagle's algorithm for all TCP connections (unless explicitly disabled via the TCP_NODELAY setsockopt())</p> <p>Notes:</p> <ul style="list-style-type: none"> ▪ z/OS implements a "Relaxed" Nagle algorithm that allows some small segments to be sent under certain conditions. ▪ This parameter is related to the DELAYACKS/ NODELAYACKS parameter setting. NODELAYACKS is another method for solving latency issues related to Nagle's algorithm on the receiving TCP end point. 	N/A	<i>NAGLE</i>

New TCPCONFIG parameters in TCPIP profile (cont)

New TCPCONFIG parameter	Function	Range	Default
QUEUEDRTT	<p>Outbound serialization is used when a high latency network is detected.</p> <ul style="list-style-type: none"> ▪ avoids out of order outbound packets ▪ high latency is defined as an RTT and SRTT of 20 ms <p>This parameter specifies the latency (in milliseconds) used for outbound serialization. Use with care, typically under the direction of IBM service</p>	0-50 ms	20 ms (same as internal setting on previous releases)
FRRTHRESHOLD	<p>Fast retransmit/fast recovery (FRR) engages after 3 duplicate ACKs are received</p> <ul style="list-style-type: none"> ▪ duplicate ACKs indicate lost or out of order packets ▪ RFC 2581 defines FRR <p>This parameter defines the number of duplicate ACKs required before FRR is engages for a TCP connection. <i>Use with care!</i></p>	1-2048 duplicate ACKs	3 duplicate ACKs (same as internal setting in previous releases)

ICMP outbound flood prevention

- Communications Server paces TCP and EE outbound traffic
 - One benefit of this is that it prevents excessively long queue buildups that could tie up large amounts of CSM storage.
 - This type of pacing doesn't apply to ICMP, RAW, or non-EE UDP traffic.
- Communications Server also limits the number of inbound packets on a single QDIO interface at a given time.
- V2R1 added similar protection for outbound ICMP, RAW, and UDP traffic by dropping outbound QDIO packets when approaching CSM constrained or critical conditions, and with the outbound QDIO queues in a congested state.



Miscellaneous items

- Trace event exit regardless of CFS trace option
 - The Coupling Facility Services (CFS) component always traces connection related activities and other important information in the mini-trace table for the Coupling Facility structures: ISTGENERIC, EZBDVIPA and EZBEPOR
- Additional diagnostic data associated with OMPROUTE heartbeats is captured to aid in the analysis of OMPROUTE responsiveness problems
- OMPROUTE messages reporting failures in adding, changing, or deleting a route contain more information on the failing route
- The FTP client has been enhanced with trace messages to assist with the diagnosis of problems opening files. In addition to the already existing EZA2564W messages documenting a failure, these trace messages provide additional information about the root cause of the failure.
- TSO/VTAM provides the ability to translate Extended English characters for the TPUT EDIT macro instruction
- Support is added for cross-memory TPUT messages from TSO/VTAM to z/OSMF ISPF address spaces.

Miscellaneous items ...

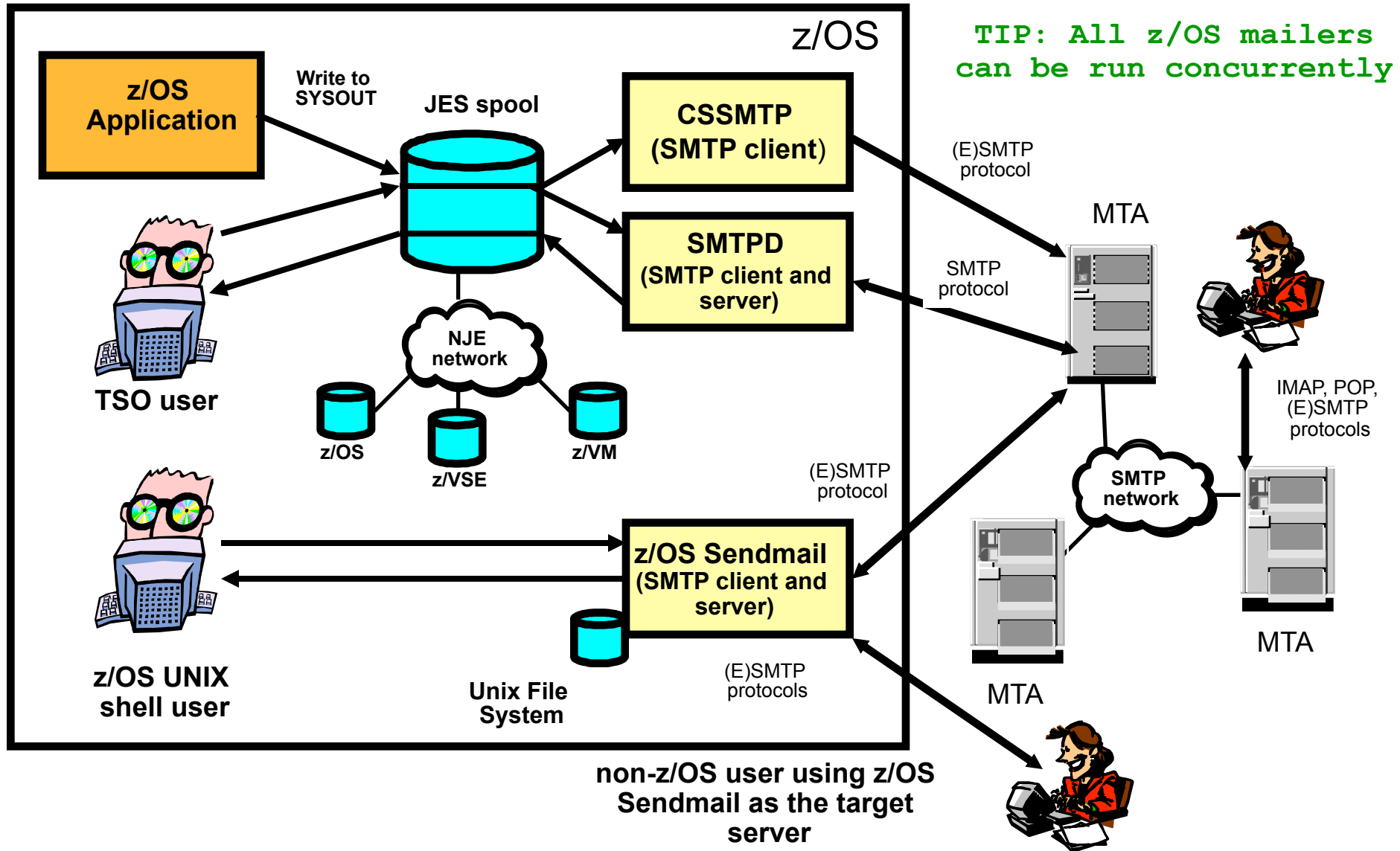
- z/OS CS provides an API for determining the SyslogD configuration file location and name.
- Enhanced IDS IP fragment detection
 - V2R1 changes the fragmentation attack probe to no longer consider fragment length as a criteria. Checks are based purely on whether overlays occur and whether they change the packet content.

z/OS V2R1 Communications Server Technical Update

**Appendix B:
Background on the Statement of
Direction for
z/OS CS Internet Mail Applications:
Sendmail and SMTPD**



z/OS Communications Server Mail Solutions – today’s picture



z/OS SMTP (MAIL) Strategy for aging components



Communications Server provides the following Simple Mail Transfer Protocol (SMTP) transports and plans to remove them in a future release:



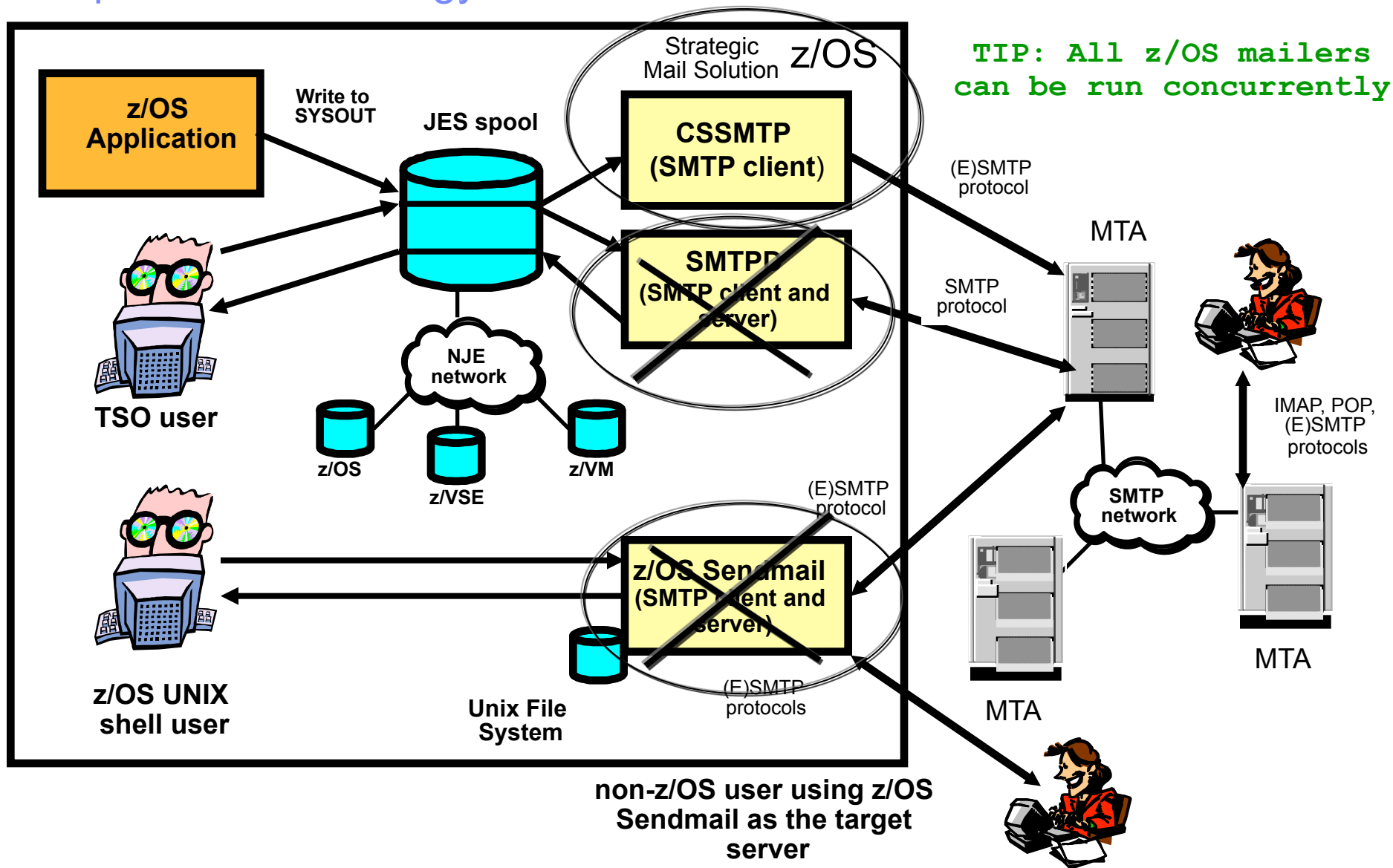
- **Sendmail** – provides both SMTP client and server roles.
 - Runs as a SMTP server (listener) on z/OS to receive mail from SMTP clients (locally or remotely) for delivery to Unix mail boxes or for forwarding.
 - Not heavily used on z/OS. Primary use case: Client program is used by tooling and applications to send mail from z/OS.

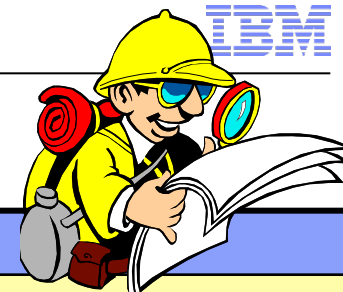
- ***SMTPD NJE Mail Gateway*** – provides both SMTP client and server roles.
 - Heavily used application by batch jobs (some TSO usage)
 - In gateway role, reads spool datasets which contain mail messages allocated by batch jobs and TSO users, then uses SMTP client capability to send mail from z/OS.
 - Also provides SMTP server (listener) role to receive mail from SMTP clients (locally or remotely - ex: typically a local sendmail client) for delivery to TSO users or for forwarding.
 - No support for IPv6 and TLS/SSL, does not perform or scale well

CSSMTP NJE Mail Gateway is the Strategic z/OS SMTP Transport



- Provided in z/OS V1R11 to address the aging SMTPD NJE Gateway.
 - Designed to provide easy migration from SMTPD for those that send mail from z/OS
 - Provides the NJE mail gateway role – removing messages from spool and transporting them using SMTP (*outbound email focus*)
 - Many customers have already migrated
 - Has significant advantages over the SMTPD NJE Gateway:
 - Performance and scalability is significantly better
 - In internal performance benchmarks (sending 4000 emails) CSSMTP was **4.5 times** faster while **using half as much CPU** than SMTPD
 - Support for latest mail standards
 - Support for IPv6 and AT-TLS
 - Uses system translation services
- Limitations
 - Does not provide an SMTP listener capability (i.e. inbound email support, including delivery of mail to TSO users)
 - Some incompatibilities with SMTPD NJE gateway
 - SMTPD NJE provided lax enforcement of some SMTP standards
 - Some emails that were accepted by SMTPD NJE get rejected by CSSMTP
 - Several problems in this area have already been corrected (i.e. allowing CSSMTP to accept and process these emails)

Proposed Mail Strategy





For more information

URL	Content
http://www.twitter.com/IBM_Commserver 	IBM z/OS Communications Server Twitter Feed
http://www.facebook.com/IBMCommserver 	IBM z/OS Communications Server Facebook Page
https://www.ibm.com/developerworks/mydeveloperworks/blogs/IBMCommserver/?lang=en	IBM z/OS Communications Server Blog
http://www.ibm.com/systems/z/	IBM System z in general
http://www.ibm.com/systems/z/hardware/networking/	IBM Mainframe System z networking
http://www.ibm.com/software/network/commserver/	IBM Software Communications Server products
http://www.ibm.com/software/network/commserver/zos/	IBM z/OS Communications Server
http://www.redbooks.ibm.com	ITSO Redbooks
http://www.ibm.com/software/network/commserver/zos/support/	IBM z/OS Communications Server technical Support – including TechNotes from service
http://www.ibm.com/support/techdocs/atmastr.nsf/Web/TechDocs	Technical support documentation from Washington Systems Center (techdocs, flashes, presentations, white papers, etc.)
http://www.rfc-editor.org/rfcsearch.html	Request For Comments (RFC)
http://www.ibm.com/systems/z/os/zos/bkserv/	IBM z/OS Internet library – PDF files of all z/OS manuals including Communications Server
http://www.ibm.com/developerworks/rfe/?PROD_ID=498	RFE Community for z/OS Communications Server
https://www.ibm.com/developerworks/rfe/execute?use_case=tutorials	RFE Community Tutorials

For pleasant reading

Please complete your session evaluation

- z/OS V2R1 Communications Server Technical Update, Part 2
- Session # 14678
- QR Code:



Find us on Facebook at
<http://www.facebook.com/IBMCommserver>



Follow us on Twitter at
http://www.twitter.com/IBM_Commserver



Read the z/OS Communications Server blog at
<http://tinyurl.com/zoscsblog>



Visit the z/OS CS YouTube channel at
<http://www.youtube.com/user/zOSCommServer>

