# z/OS® V1R10 Communications Server

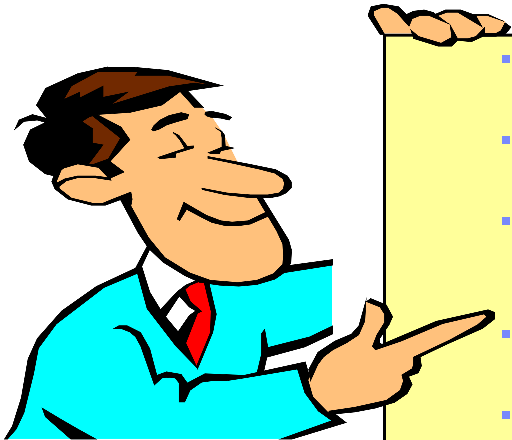
## *Overview : Platform enhancements*

@business on demand software



This presentation is an overview of platform enhancements in Communications Server for z/OS V1R10.

## Agenda



- What is z/OS Communication Server?
- TCP/IP performance Improvements
- IBM System z10 EC™ Hipersockets Multiple Write Facility
- Multiple VLANs per OSA per stack per IP protocol version
- INTERFACE statement support for IPv4 QDIO
- Resolver support for EDNS0

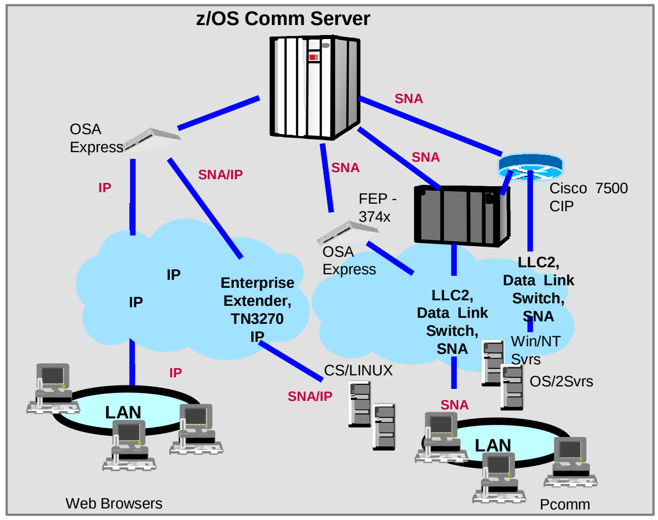



This presentation reviews the function of the z/OS Communications Server.

The platform enhancements for CS z/OS V1R10 include TCP/IP performance improvements, IBM System z10 EC Hipersockets Multiple Write Facility support, Multiple Virtual Local Area Networks (VLAN) per Open Systems Adapter (OSA) per stack per IP protocol version, INTERFACE statement support for IPv4 QDIO, and Resolver support for EDNS0.

## Overview: What is z/OS Communications Server?

- **A z/OS base element**
- **Supports networking throughout an enterprise**
  - Secure TCP/IP, SNA, and UNIX® protocols
- **Consists of two components**
  - IP Services
  - SNA Services

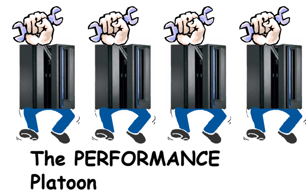


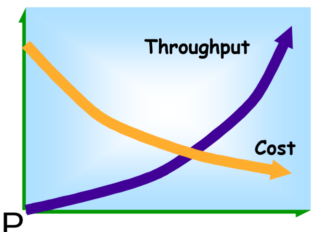

 

The z/OS Communications Server is a z/OS base element consisting of IP and SNA services that support networking throughout an enterprise.

# TCP/IP performance improvements

- TCP/IP path length reduction for primarily Request/Response workload
- TCP/IP data path ECSA relief
- Cache line contention (also known as hot cache line)
- IPSec use of CP Assist for Cryptographic Function (CPACF) directly when possible
- Reduce ECSA storage for TN3270-related VTAM® control blocks
- Enhance performance for multithreaded UDP applications



For the inbound TCP/IP datapath, Communications Server now keeps message headers in CSM storage, avoiding copies to Extended Control Storage Area (ECSA).

Cache line contention is a CPU scalability issue. Cache line contention occurs when many CPUs attempt to update the same cache line (storage), for example, as the result of Compare and Swap (CS) operations on a shared control block by many concurrently processing threads (subtasks) each dispatched on different CPUs. When this happens, cache memory needs to be invalidated, CPU processing is halted, cache rebuilt, and CPU resumes. The performance impact of cache line contention increases as number of CPUs increases. z/OS CS is known to cause cache line contention under certain conditions, and was examined and changes were made where possible to reduce the impacts of cache line contention.

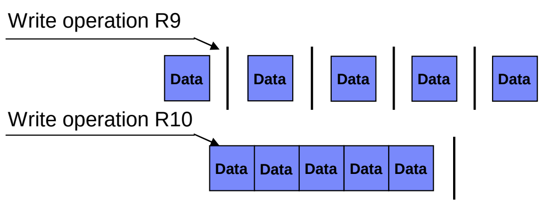
Before V1R10, IPSec uses Integrated Cryptographic Service Facility (ICSF) for all crypto functions, which is convenient from an implementation perspective. In V1R10, IPSec will save a few cycles by using the CPACF, CP Assist for Cryptographic Function, directly, for the cryptographic functions that are supported by the CPACF. Throughput improvements are greatest for especially short datagrams (< 1K). Anything else that currently is directed to ICSF will continue to be directed to ICSF.

Each SNA session between the TN3270 server and SNA applications use a Request Parameter List (RPL) that is located in TN3270 server private storage. Most of the time, these RPLs are copied to ECSA as CRPL control blocks, and Each CRPL takes up 160 bytes of storage. This means, for example, that for 6,000 TN3270 sessions this is about 1 MB of ECSA storage, for 60,000 TN3270 sessions this is about 10 MB of ECSA storage, and for 256,000 TN3270 sessions this is about 40 MB. z/OS V1R10 has changed TN3270 and VTAM processing so VTAM will be able to use the RPL as-is without copying it to ECSA.

Previous to V1R10, all threads with an outstanding read on a UDP socket are posted when data arrives. Sometimes a race condition occurs: one thread gets the data and all the others (while executing in the UDP layer) need to re-suspend waiting for more datagrams to arrive. The problem gets worse as the number of threads increases. This can make a multithreaded UDP receiver design perform worse than a single-threaded design. In z/OS V1R10, only one thread will be posted when data arrives.

# IBM System z10 EC Hipersockets Multiple Write Facility

- Hipersockets can now move multiple output data buffers
  in one write operation
  - Reduces CPU utilization
  - For large outbound messages
    - Used when message spans Hipersocket frame size
  - For more CPU savings, the processing for these large writes can be offloaded to a zIIP.




5 Overview : Platform enhancements 

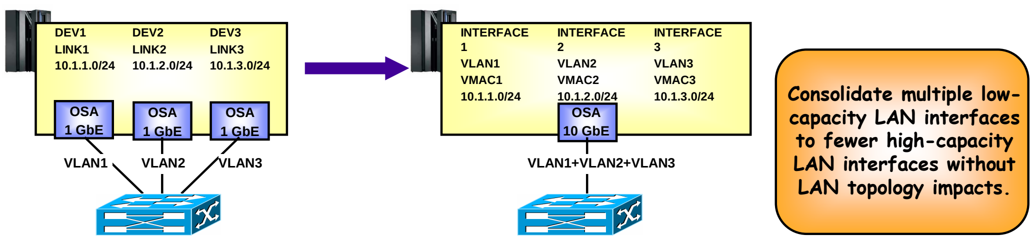
The newly announced IBM System z10 includes a new function called Hipersockets Multiple Write. This allows multiple data buffers to be moved from one system image to another across Hipersockets with one operation. This can reduce CPU utilization.

When enabled, HiperSockets Multiple Write will be used anytime a message spans the Hipersockets frame size, thus requiring multiple output buffers to transfer the message. Therefore, it will only be used for larger outbound messages. Spanning multiple output data buffers can be affected by several factors including the Hipersockets frame size, Application socket send size, TCP send size and MTU size.

IBM

# Multiple VLANs per OSA per stack per IP protocol version

- Current TCP/IP stack support
  - Two VLANs per OSA port
  - One each for IPv4 and IPv6
- With consolidation to higher bandwidth OSA-Express (for example 10 Gig)
  - Too restrictive
- Support added for up to eight VLANs per IP protocol per OSA port for one TCP/IP stack




6 Overview : Platform enhancements 

The previous limitation of one VLAN per IP protocol was too restrictive because it was not possible to retain existing network interface and IP subnet topology when consolidating multiple LANs to one LAN. This would require IP renumbering.

With the new support, Each VLAN on the same OSA port must use unique, non-overlapping IP subnets or prefixes . This will be enforced by the TCP/IP stack.

Each VLAN must be defined using a new IPv4-enabled version of the INTERFACE configuration statement, which only supports QDIO interfaces.

Each VLAN must use layer-3 virtual MAC addresses and each VLAN must have a unique virtual MAC address.

# INTERFACE statement support for IPv4 QDIO interfaces

- In z/OS V1R4,
  - Added new INTERFACE TCP/IP PROFILE statement
  - Only for IPv6 network interfaces
- z/OS V1R10 extends use of the INTERFACE statement
  - Now for IPv4 QDIO network interfaces
  - Allows for consistency, subnet definition, SOURCEVIPA
- Example:

```
INTERFACE OSAQDIO24
    DEFINE IPAQENET
    PORTNAME OSAQDIO2
    SOURCEVIPAINT VIPAV4
    IPADDR 10.1.1.1/24
    MTU 8992
    VLANID 200
    VMAC ROUTEALL
```

SOURCEVIPA specification

Subnet mask can be specified

Much simplified configuration of IPv4 network interfaces




When the INTERFACE statement was introduced for IPv6, IPv4 interfaces continued to require use of DEVICE, LINK, HOME, and optionally BSDROUTINGPARMS statements to define all the attributes of an IPv4 network interface.

This new statement enables enforcement of unique subnet masks for multiple VLANs on same OSA port. It was only added for IQDIO interfaces. Non-QDIO IPv4 network interfaces continue to require the old configuration syntax (DEVICE/LINK).

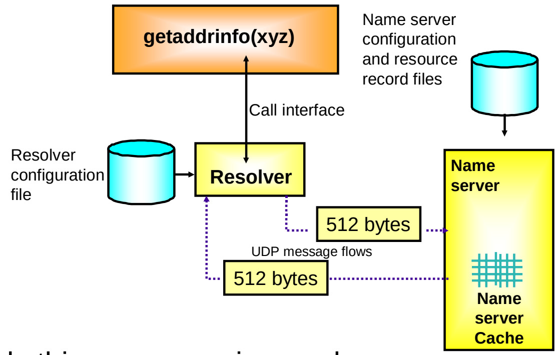
With the DEVICE/LINK method of defining IPv4 interfaces, the SOURCEVIPA address to be used had to be inferred from placement in the home list. For this new statement, you can specify the specific SOURCEVIPA to indicate which static VIPA IP address will be used as the source on outbound connections are established over this interface (unless overridden by more specific source IP addressing rules).

You can also specify a subnet mask by specifying a non-zero value as prefix length on the IPADDR keyword (in this example a prefix length of 24 maps to 255.255.255.0). When you do this,  the TCP/IP stack will inform OSA to only perform ARP processing for a VIPA if the VIPA is configured in the same subnet as the OSA (as defined by this subnet mask).

You do not have to code a MAC address. These definitions will allow the OSA adapter to generate a unique MAC address for this interface.

# Resolver support for  EDNS0

- Original DNS standards limited the UDP message size between a resolver and the name server to 512 bytes.
  - For IPv6 or IPv4 and IPv6 mixed environments, the 512-byte message size has become a severe limitation




- An updated DNS standard extends this message size, and z/OS CS V1R10 resolver adds support for this version of the DNS standards
  - Support of EDNS0 is a requirement for IPv6 compliance and is part of a set of standards that IPv6 installations require support for.

DNS RFC currency. IPv6 compliance.




Under the original  DNS standard, if DNS could not respond back in a 512 byte message, it would be indicated in the reply and the resolver is then to repeat the query using a TCP connection to the name server instead. This can be costly in terms of elapsed time, but for IPv4-only environments, this very seldom happened so the 512-byte message size was not a severe limitation.

The updated DNS standards are described in RFCs 2671 and 3226. These are generally referred to as EDNS0 (Extension mechanisms for DNS version 0).

The z/OS CS BIND9 DNS server already supports EDNS0 with a maximum message size of 4096 bytes.

# Feedback

## Your feedback is valuable

You can help improve the quality of IBM Education Assistant content to better meet your needs by providing feedback.

- Did you find this module useful?
- Did it help you solve a problem or answer a question?
- Do you have suggestions for improvements?

Click to send e-mail feedback:

mailto:iea@us.ibm.com?subject=Feedback_about_wnplate.ppt

This module is also available in PDF format at: ../wnplate.pdf




You can help improve the quality of IBM Education Assistant content by providing feedback.

# Trademarks, copyrights, and disclaimers

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

| | | | | | |
|---|---|---|---|---|---|
| IBM | System z | System z10 | VTAM | z/OS | z10 EC |

A current list of other IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

UNIX is a registered trademark of The Open Group in the United States and other countries.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements or changes in the products or programs described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

Information is provided "AS IS" without warranty of any kind. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (for example, IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products.

IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance 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 or performance improvements equivalent to the ratios stated here.



Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.