Communications Server z/OS V1R5 and V1R6 Technical Update

zSeries[7] Hardware Exploitation For V1R5

© Copyright International Business Machines Corporation 2004. All rights reserved.

Topics

- z/OS V1R5
  - OSA performance enhancements
    - –Read storage control
    - –Inbound performance
    - –Offload checksum processing
  - Full VLAN support
  - Expanded QDIO and iQDIO capacity limits
  - HiperSockets broadcast support

© Copyright International Business Machines Corporation 2004. All rights reserved.

z/OS V1R5 OSA Performance Enhancements

Copyright International Business Machines Corporation 2004. All rights reserved.

OSA-Express performance enhancements Overview

OSA-Express performance enhancements provided in z/OS V1R5:

- **Read Storage Control**
  - Provide more granularity and control of fixed storage allocation per OSA
- **Inbound Performance**
  - Provide better control for optimizing CPU utilization vs. latency
- **Checksum Offload**
  - Provide for CPU reduction when checksum can be offloaded
  - Checksum processing can consume a significant percentage of pathlength in the TCP/IP stack.

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO read storage use control

- Each OSA-Express QDIO device and HiperSockets device requires a lot of fixed storage for read processing
- VTAM provides start options to configure the amount of read storage but these settings apply globally
  - QDIOSTG applies to all OSA-Express QDIO devices
  - IQDIOSTG applies to all HiperSockets devices
- Prior to z/OS V1R5, there was no way to configure the storage usage for a specific device
- z/OS V1R5 provides new keywords in the TCP/IP profile to override the global VTAM read storage setting for a specific OSA-Express QDIO or HiperSockets device
- Can specify one of the following values:
  - **GLOBAL**
    The amount of storage is determined by the QDIOSTG or IQDIOSTG VTAM start option. This is the default.
  - **MAX**
    You expect a heavy inbound workload over this adapter
  - **AVG**
    You expect a medium inbound workload over this adapter
  - **MIN**
    You expect a light inbound workload over this adapter
- For HiperSockets, only affects devices with 64K frame size

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO read storage control - configuration

- New keyword on
  - LINK statement for IPAQENET, IPAQTR, and IPAQIDIO and
  - INTERFACE statement for IPAQENET6

```
>>-LINK--link_name--IPAQENET--device_name--+---------+--+---------------+---->
                                           '-IPBCAST-'  '-VLANID id-----'

  .-READSTORAGE-----GLOBAL-.   .-INBPERF-----BALANCED---.
>-+------------------------+---+-----------------------+--|
  +-READSTORAGE--+--MAX----+   +-INBPERF--+--MINCPU-----+
                 +--AVG----+              '--MINLATENCY-'
                 '--MIN----'

    .-IFSPEED 100000000-.
>--+-------------------+-------------------------------------><
   +-IFSPEED ifspeed---+
   '-IFHSPEED ifhspeed-'
```

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO read storage control - Netstat

- Netstat DEVLINKS/-d enhanced to display the READSTORAGE setting and the amount of read storage

```
DevName: OSAQDIO4          DevType: MPCIPA
  DevStatus: Ready
  LnkName: OSAQDIOLINK       LnkType: IPAQENET   LnkStatus: Ready
    NetNum:  0    QueSize: 0    Speed: 0000000100
    IpBroadcastCapability: No
    CfgRouter: Non                     ActRouter: Non
    ArpOffload: Yes                    ArpOffloadInfo: Yes
    ActMtu:  1492
    VLANid:  1260                      VLANpriority: Enabled
    ReadStorage: GLOBAL (8064K)        InbPerf: Balanced
    ChecksumOffload: Yes
```

- New SNMP MIB object
  - ibmMvsLinkReadStorageSize - the amount of fixed storage in kilobytes for read processing (for OSA-Express QDIO and HiperSockets)

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO read storage control - Display TRLE

- The DISPLAY TRL,TRLE= command output was enhanced to display the amount of read storage for each active OSA-Express QDIO or HiperSockets TRLE

```
IST075I NAME = OSAQ4, TYPE = TRLE
IST1954I TRL MAJOR NODE = OSAQDIO2
IST486I STATUS= ACTIV, DESIRED STATE= ACTIV
IST087I TYPE = LEASED             , CONTROL = MPC , HPDT = YES
IST1715I MPCLEVEL = QDIO       MPCUSAGE = SHARE
IST1716I PORTNAME = OSAQDIO4   LINKNUM =   0   OSA CODE LEVEL = 0330
IST1577I HEADER SIZE = 4096 DATA SIZE = 0 STORAGE = ***NA***
IST1221I WRITE DEV = 2E01 STATUS = ACTIVE     STATE = ONLINE
IST1577I HEADER SIZE = 4092 DATA SIZE = 0 STORAGE = ***NA***
IST1221I READ  DEV = 2E00 STATUS = ACTIVE     STATE = ONLINE
IST1221I DATA  DEV = 2E02 STATUS = ACTIVE     STATE = N/A
IST1724I I/O TRACE = OFF  TRACE LENGTH = *NA*
IST1717I ULPID = TCPCS
IST1915I IQDIO ROUTING DISABLED
IST1910I READ STORAGE    4.0M(64 SBALS)
IST1757I PRIORITY1: UNCONGESTED PRIORITY2: UNCONGESTED
IST1757I PRIORITY3: UNCONGESTED PRIORITY4: UNCONGESTED
..
```

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO read storage control - notes

NOTES

- You can use the VTAM Tuning Statistics for guidance on what to configure for READSTORAGE. See the documentation on the QDIOSTG and IQDIOSTG VTAM start options in the SNA Resource Definition Reference for more details. Note: Comm Server uses CSM dataspace buffers backed by 64-bit real for this read storage.

OSA-Express QDIO

READSTORAGE value	Storage
MAX	4 MB
AVG	2 MB
MIN	1 MB
GLOBAL	based on QDIOSTG VTAM start option

HiperSockets with 64K frame size

READSTORAGE value	Storage
MAX	7.8 MB
AVG	6 MB
MIN	4 MB
GLOBAL	based on IQDIOSTG VTAM start option

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO read storage control - details

- You only need to take action if you want to override the global VTAM default settings for a specific OSA-Express QDIO or HiperSockets device.
- There is no way to override the default for the HiperSockets device used for DYNAMICXCF (IUTIQDIO). You can, however, use the IQDIOSTG VTAM start option to set a value for the IUTIQDIO device and use the READSTORAGE setting to specify different values for other HiperSockets devices if desired.
- If you use the same OSA-Express for both IPv4 and IPv6 traffic, you need to specify the same READSTORAGE setting on both the corresponding LINK and INTERFACE statements

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO inbound performance

- The performance of an OSA-Express QDIO device is impacted by how frequently the OSA interrupts the host to process inbound packets
  - More frequent interruptions lead to minimized latency but increased CPU consumption
  - Less frequent interruptions lead to decreased CPU consumption but increased latency
- Prior to z/OS V1R5, there was no way to configure the desired inbound performance characteristics for a specific device
- Provide new keyword in TCP/IP profile to specify the desired inbound performance behavior from an OSA-Express in QDIO mode
  - PTFed back to z/OS V1R4 via APAR PQ92262
- Can specify one of the following values:
  - **MINCPU** - instructs the adapter to minimize host interrupts, thereby minimizing host CPU consumption. This mode of operation may result in minor queuing delays for packets into the host, and is not recommended for workloads with demanding latency requirements.
  - **MINLATENCY** - instructs the adapter to minimize latency, by immediately presenting received packets to the host. This mode of operation will generally result in higher CPU consumption than the other two settings, and is recommended only for workloads with demanding latency requirements. This setting should only be used if host CPU consumption is not an issue.
  - **BALANCED** (default) - instructs the adapter to strike a balance between MINCPU and MINLATENCY

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO inbound performance - Notes

NOTES

- The frequency with which an OSA-Express in QDIO mode interrupts the host for inbound data has evolved as follows:
  - Stage 1: OSA determined the frequency based on the number of read buffers available.
  - Stage 2: OSA determined the frequency based on fixed values set by z/OS Communications Server. This function requires the same OSA microcode levels mentioned below. For pre-V1R5, this function also requires one of the following VTAM APARs. See one of these APARs for more details. Note: These values correspond to the default BALANCED value with this new lineitem.
    - OS/390 V2R10 OA02011/UA01438
    - z/OS V1R2 OW56896/UA00067
    - z/OS V1R4 OW56019/UW94043
  - Stage 3: OSA determines the frequency based on values set by z/OS Communications Server. The difference from stage 2 is that in V1R5, the values are based on the desired performance behavior specification in the TCP/IP profile and, therefore, are configurable in a general sense (whereas the values for stage 2 were fixed and not at all configurable). This function requires the same OSA microcode levels as stage 2.
- If you use the same OSA-Express for both IPv4 and IPv6 traffic, you need to specify the same INBPERF setting on both the corresponding LINK and INTERFACE statements
- INBPERF will have no effect if OSA-Express microcode is downlevel
- The z/OS V1R5 support requires one of the following OSA Express microcode levels - or newer:

Processor	Microcode level
G5/G6	4.28
zSeries 2064 GA2	2.29
zSeries 2064 GA3	3.23

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO inbound performance - Configuration

- New keyword on
  - LINK statement for IPAQENET and IPAQTR
  - INTERFACE statement for IPAQENET6

```
>>-LINK--link_name--IPAQENET--device_name--+---------+--+---------------+---->
                                           '-IPBCAST-'  '-VLANID id-----'

  .-READSTORAGE-----GLOBAL-.   .-INBPERF-----BALANCED---.
>-+------------------------+---+-----------------------+--|
  +-READSTORAGE--+--MAX----+   +-INBPERF--+--MINCPU-----+
                 +--AVG----+               '--MINLATENCY-'
                 '--MIN----'

    .-IFSPEED 100000000-.
>--+-------------------+-------------------------------------><
   +-IFSPEED ifspeed---+
   '-IFHSPEED ifhspeed-'
```

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO inbound performance - Netstat

- Netstat DEVLINKS/-d enhanced to display the INBPERF setting

```
DevName: OSAQDIO4          DevType: MPCIPA
  DevStatus: Ready
LnkName: OSAQDIOLINK       LnkType: IPAQENET   LnkStatus: Ready
    NetNum:  0   QueSize: 0   Speed: 0000000100
    IpBroadcastCapability: No
    CfgRouter: Non                   ActRouter: Non
    ArpOffload: Yes                  ArpOffloadInfo: Yes
    ActMtu:  1492
    VLANid:  1260                    VLANpriority: Enabled
    ReadStorage: GLOBAL (8064K)      InbPerf: Balanced
    ChecksumOffload: Yes
```

- New SNMP MIB object
  - ibmMvsLinkInboundPerfType - indicates the INBPERF setting (for OSA-Express QDIO)

© Copyright International Business Machines Corporation 2004. All rights reserved.

Performance impacts of inbound performance control

- Inbound Performance control:
  - New parm on Link statement : INBPERF
  - BALANCED : Default (Recommended)
  - MINCPU : minimizes cpu
    - TPUT: - 29.6 to - 0.5 % CPU/Tran: - 5.3 to + 0.7 %
    - Reduces cpu/tran for RR & CRR with a loss of tput
  - MINLATENCY: minimizes latency
    - TPUT: - 2.1 to + 5.2 % CPU/Tran: - 4.8 to + 2.4 %
    - Improves tput for RR and CRR workloads,
    - Degrades Streams workload

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO checksum offload

- Offload most IPv4 checksum processing to OSA-Express in QDIO mode
- Provide improved performance for IPv4 traffic
- Supported on the following OSA-Express features: (which require an IBM eServer zSeries 890 or 990)
  - Feature # 1364 GbE LX
  - Feature # 1365 GbE SX
  - Feature # 1366 1000BASE-T Ethernet when configured to operate at 1 Gbps
- Only applies to IPv4 packets
- Only applies to packets which go onto the LAN
  - Not to packets that are passed back into the zSeries to an LPAR sharing the OSA-Express LAN port
- Applies to TCP, UDP, and IP header checksums
- Applies to both inbound and outbound
- Exceptions
  - Fragmentation/reassembly
  - IPSEC
  - Packets between 2 stacks sharing the OSA
  - Outbound multicast and broadcast
  - Some outbound TCP control packets (e.g. SYN, RST)

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO checksum offload - Configuration and Netstat reports

- No configuration is needed to enable the function
- Netstat DEVLINKS/-d enhanced to display whether the OSA supports checksum offload

```
DevName: OSAQDIO4          DevType: MPCIPA
  DevStatus: Ready
  LnkName: OSAQDIOLINK       LnkType: IPAQENET   LnkStatus: Ready
    NetNum:  0    QueSize: 0    Speed: 0000000100
    IpBroadcastCapability: No
    CfgRouter: Non                    ActRouter: Non
    ArpOffload: Yes                   ArpOffloadInfo: Yes
    ActMtu:  1492
    VLANid:  1260                     VLANpriority: Enabled
    ReadStorage: GLOBAL (8064K)       InbPerf: Balanced
    ChecksumOffload: Yes
```

- New SNMP MIB object
  - ibmMvsLinkChecksumOffloadEnabled - indicates whether the adapter is enabled for checksum offload (for OSA-Express QDIO)

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO checksum offload - notes

NOTES

- Need new OSA-Express feature (and therefore a zSeries 890 or 990) to get the new function
- If OSA detects a checksum failure on inbound, OSA will still present the packet to the stack. This way, the checksum error discard will still appear in the TCP/IP CTRACE for serviceability purposes.
- TCP/IP will perform checksum for all packets to any devices which do not support the checksum offload function

© Copyright International Business Machines Corporation 2004. All rights reserved.

Performance impacts of checksum offload

- Checksum Offload
  - TCP/IP stack Checksum function offloaded to OSAE adapter
  - Requires a zSeries 990 or 890 system
  - Request/Response Workload :
    - TPUT: Equivalent
      - CPU/Tran: - 0.6 to - 1.4 %
  - Connect Request/Response Workload :
    - TPUT: - 2.1 %
      - CPU/Tran: - 5.5 to - 6.0 %
  - Streams Workload :
    - (TCP) TPUT: - 1.6 %
      - CPU/Tran: - 11.5 to - 13.8 %
    - (EE) TPUT: + 145 %
      - CPU/Tran: - 9.6 to - 14.3 %

© Copyright International Business Machines Corporation 2004. All rights reserved.

Full VLAN support for OSA-Express QDIO in z/OS V1R5

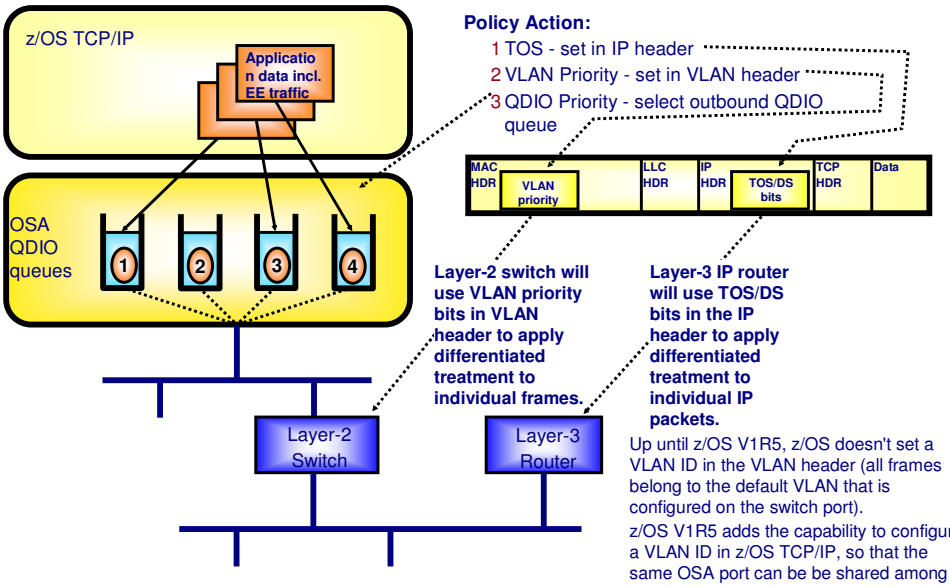
Copyright International Business Machines Corporation 2004. All rights reserved.

VLANs

- Virtual Local Area Network (VLAN) technology is becoming more important to network planning for many customers
- LAN - Broadcast domain
  - Nodes on LAN can communicate with each other without a router
  - Router needed between LANs
- VLAN - A configured logical grouping of nodes using switches
  - Nodes on VLAN can communicate as if they were on same LAN
  - Router needed between VLANs
- z/OS customers want to take advantage of the benefits of VLAN
  - Can improve network performance by reducing traffic on a physical LAN
  - Can enhance security by isolating traffic
  - Provides more flexibility in configuring networks
- z/OS Communications Server V1R2 provided VLAN priority tagging (limited to the null VLAN)
- IEEE802.1p: Priority tagging
- IEEE802.1q: VLAN ID tagging

© Copyright International Business Machines Corporation 2004. All rights reserved.

How are QDIO priority, VLAN priority, and TOS/DS bit settings used by the networking components?



Up until z/OS V1R5, z/OS doesn't set a VLAN ID in the VLAN header (all frames belong to the default VLAN that is configured on the switch port).

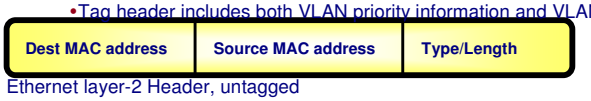
z/OS V1R5 adds the capability to configure a VLAN ID in z/OS TCP/IP, so that the same OSA port can be be shared among LPARs that are connected to different VLANs on the same OSA port.

© Copyright International Business Machines Corporation 2004. All rights reserved.

VLAN tagging basics

Two types of frames in a VLAN environment:

- Untagged frame
  - –No tag header following the source MAC address
- Tagged frame
  - –Priority-tagged frame
    - •Tag header includes only VLAN priority information, but no VLAN ID (VLAN ID is zero and is referred to as a null-tagged frame)
  - –VLAN-tagged frame
    - •Tag header includes both VLAN priority information and VLAN ID



Ethernet layer-2 Header, untagged

Ethernet layer-2 Header, tagged

Tag Control Information

© Copyright International Business Machines Corporation 2004. All rights reserved.

Full VLAN support in z/OS V1R5

- Provides full VLAN support for OSA-Express QDIO by allowing VLAN ID to be configured
- Supported on these OSA-Express features: (which require an IBM eServer zSeries 890 or 990)
  - Gigabit Ethernet (feature #s 2364, 2365, 1364, 1365)
  - Fast Ethernet (feature # 2366)
  - 1000BASE-T Ethernet (feature # 1366)
- Also supported on the following OSA-Express features (at system driver level 3G) on an IBM eServer zSeries 800 or 900:
  - Gigabit Ethernet
  - Fast Ethernet
- Conforms to the IEEE 802.1Q standard
  - VLAN tag in MAC header contains VLAN ID
- Allows VLAN priority tagging to be used with VLAN ID
- Configure VLAN ID in TCP/IP profile
  - Each stack can only specify one VLAN ID for each OSA-Express port per IP version
    - A stack can specify separate VLAN IDs for IPv4 and IPv6 for the same OSA-Express port
  - Each stack sharing an OSA-Express port may specify a different VLAN ID
- z/OS V1R4 PTF coming (APAR PQ86508)
- Multiple OSA PRIRouters supported (PRIRouter per global VLANID support)

© Copyright International Business Machines Corporation 2004. All rights reserved.

z/OS TCP/IP LINK or INTERFACE VLAN ID specification: Global VLAN ID

- A z/OS or z/VM TCP/IP stack supports one VLAN ID (per IP protocol version) per OSA port [on the rest of this chart, we will for simplicity reasons limit ourselves to IPv4 use of VLANs).
  - This VLAN ID is configured on the LINK statement for IPv4 (or on the INTERFACE statement for IPv6).
  - This VLAN ID is referred to as the Global VLAN ID.
  - All IP packets sent over that OSA port from this TCP/IP stack will be tagged with this VLAN ID in the frame header by the OSA Express adapter.
  - All inbound broadcast and multicast IP packets will be handed up to this TCP/IP stack only if the VLAN ID in the frame matches the Global VLAN ID registered by the stack.
  - All inbound unicast IP packets will be handed up to this TCP/IP stack only if the VLAN ID in the frame header macthes the Global VLAN ID registered by this stack - and -
    - the destination IP address matches one of the HOME IP addresses that were registered in the adapter by this TCP/IP stack

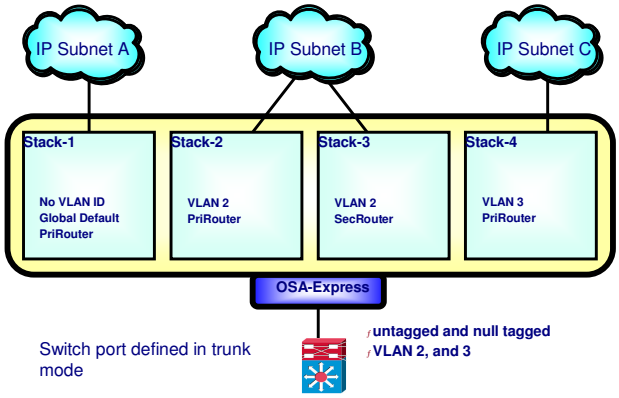
      --- OR ---
    - the destination IP address in the packet is not registered by any stack sharing this OSA adapter port and this stack is a PRI/SECRouter stack
- A Linux on zSeries TCP/IP stack is able to tag individual IP packets sent over the OSA port with different VLAN IDs.
  - Packets can be sent by Linux using different VLAN IDs
  - This kind of operation complicates configuration and inbound processing significantly and is not supported by either z/OS or z/VM
    - If a z/OS TCP/IP stack needs access to multiple VLANs, multiple OSA adapters are required
- If an interface does not have a global VLAN ID configured, VLAN tagging of outbound frames depends on use of VLAN priorities or not:
  - No SetSubnetPrioTosMask definitions active via the policy agent: no VLAN tags will be added
  - SetSubNetPrioTosMask definitions active via the policy agent: VLAN tags with a null VLAN ID will be added

© Copyright International Business Machines Corporation 2004. All rights reserved.

VLANs and OSA PRI/SEC Router support

- The VLANID parameter of the LINK and INTERFACE statements interacts with the PRIRouter and SECRouter parameters on the DEVICE and INTERFACE statements.
  - If you configure both a VLANID and either PRIRouther or SECRouter, then this TCP/IP instance will act as a router for this VLAN only. Frames that are received at this device for an unknown IP address will only be routed to this TCP/IP instance if they are VLAN tagged with this VLAN ID.
  - If you do not configure a VLAN ID, but do configure PRIRouter or SECRouter, then this TCP/IP instance will act as a "Global default Pri/SecRouter" for all inbound unicast frames with an unknown destination IP address

- Stack-1 acts as PriRouter for:
  - untagged frames
  - frames tagged with a null VLAN ID
  - frames tagged with an unregistered (anything but VLAN 2 or 3) VLAN ID
- Stack-2 acts as PriRouter for frames tagged with VLAN 2 while Stack-3 acts as SecRouter for that same VLAN.
- Stack-4 acts as PriRouter for frames tagged with VLAN 3



© Copyright International Business Machines Corporation 2004. All rights reserved.

Full VLAN support - Configuration

- New keyword on
  - LINK statement for IPAQENET
  - INTERFACE statement for IPAQENET6
- VLANID can be in range 1-4094

```
>>-LINK--link_name--IPAQENET--device_name--+----------+--+---------------+---->
                                           '-IPBCAST-'  '-VLANID id-----'

  .-READSTORAGE-----GLOBAL-.    .-INBPERF-----BALANCED---.
>-+------------------------+---+------------------------+--|
  +-READSTORAGE--+--MAX----+   +-INBPERF--+--MINCPU-----+
                 +--AVG----+               '--MINLATENCY-'
                 '--MIN----'

    .-IFSPEED 100000000-.
>--+-------------------+---------------------------------------><
   +-IFSPEED ifspeed---+
   '-IFHSPEED ifhspeed-'
```

© Copyright International Business Machines Corporation 2004. All rights reserved.

Full VLAN support - Netstat

- Netstat DEVLINKS/-d enhanced to display the VLAN ID and whether VLAN priority is enabled

```
DevName: OSAQDIO4           DevType: MPCIPA
  DevStatus: Ready
  LnkName: OSAQDIOLINK        LnkType: IPAQENET   LnkStatus: Ready
    NetNum:  0    QueSize: 0    Speed: 0000000100
    IpBroadcastCapability: No
    CfgRouter: Non                     ActRouter: Non
    ArpOffload: Yes                    ArpOffloadInfo: Yes
    ActMtu:  1492
    VLANid:  1260                      VLANpriority: Enabled
    ReadStorage: GLOBAL (8064K)        InbPerf: Balanced
    ChecksumOffload: Yes
```

- New SNMP MIB objects in ibmTcpipMvsLinkTable
  - ibmMvsLinkVlanId - indicates the VLAN ID configured for this interface
  - ibmMvsLinkVlanPriorityEnabled - indicates that VLAN priority tagging is being done for this interface

© Copyright International Business Machines Corporation 2004. All rights reserved.

Switch configuration - trunk or access mode

- Trunk mode (new z/OS Communications Server V1R5 support)
  - Indicates that the switch should allow all VLAN ID tagged packets to pass through the switch port without altering the VLAN ID. Trunk mode is intended for servers that are VLAN capable, and filters and processes all VLAN ID tagged packets. In trunk mode, the switch expects to see VLAN ID tagged packets inbound to the switch port.
    - Configure switch in trunk mode
    - Specify VLAN ID in TCP/IP profile
    - VLAN tagging and filtering performed by OSA
    - Each stack sharing the OSA may be on a different VLAN
    - Can use different VLAN IDs for IPv4 and IPv6
    - Can use VLAN priority tagging with VLAN ID
- Access mode (support available pre-V1R5)
  - Indicates that the switch should filter on specific VLAN IDs and only allow packets that match the configured VLAN IDs to pass through the switch port. The VLAN ID is then removed from the packet before it is sent to the server (that is, VLAN ID filtering is controlled by the switch). In access mode, the switch expects to see packets without VLAN ID tags inbound to the switch port.
    - Configure switch in access mode
    - Specify VLAN ID in access mode port of switch
    - VLAN tagging and filtering performed by switch
    - Each stack sharing the OSA must be on same VLAN
    - Must use same VLAN ID for IPv4 and IPv6

© Copyright International Business Machines Corporation 2004. All rights reserved.

General VLAN network design considerations

1 When using VLAN IDs in any TCP/IP stack sharing an OSA port, the switch port to which the OSA port is attached should be configured in trunk mode.

2 When not using VLAN IDs in any TCP/IP stack sharing an OSA port, the switch port to which the OSA port is attached should be configured in access mode.

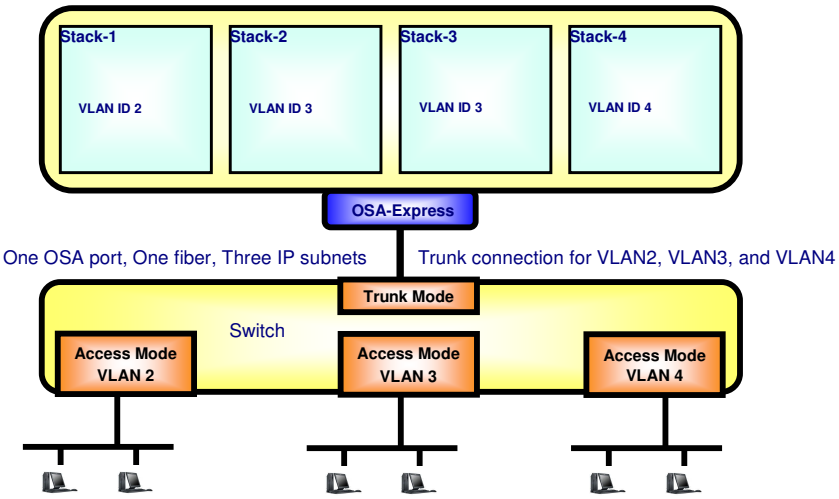
3 When a TCP/IP stack uses multiple OSA ports all to the same LAN, and a VLAN ID is used on one of those ports, VLAN IDs should be used on all ports to that same LAN.

4 Some switch vendors use VLAN ID 1 for special purposes. VLAN ID 1 should be avoided when designing VLAN-based networks.

... and please make sure your networking staff creates and maintains both a physical network diagram and a logical network diagram - they look very different when you work with VLAN configurations.

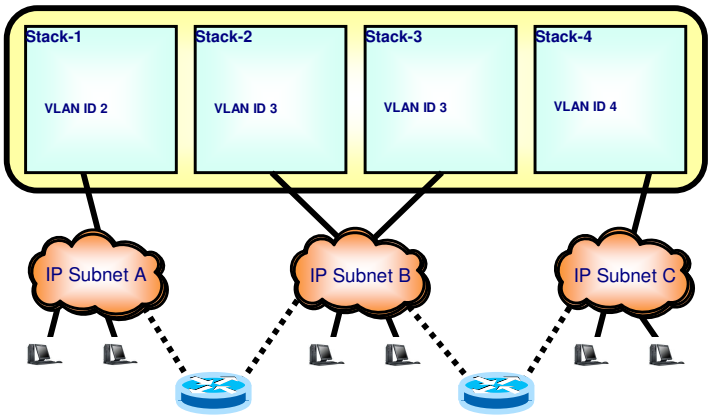
© Copyright International Business Machines Corporation 2004. All rights reserved.

Physical network connectivity diagram - Multiple stacks, Separate subnets, single OSA port



© Copyright International Business Machines Corporation 2004. All rights reserved.

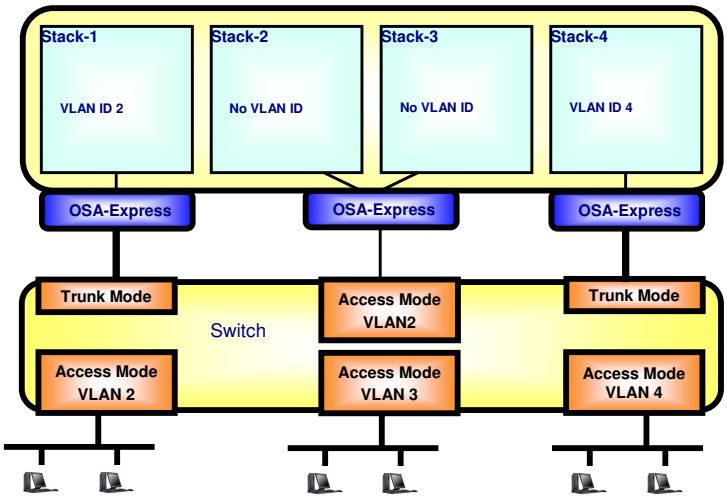
Logical network diagram - Multiple stacks, Separate subnets, single OSA port



- Depending on switch configuration, the switch may interconnect the VLANs using a layer-3 IP router function.
- The subnets may belong to different routing domains or OSPF areas.
  - Test, production, demo
- The subnets may belong to different security zones.
  - Intranet, DMZ

© Copyright International Business Machines Corporation 2004. All rights reserved.

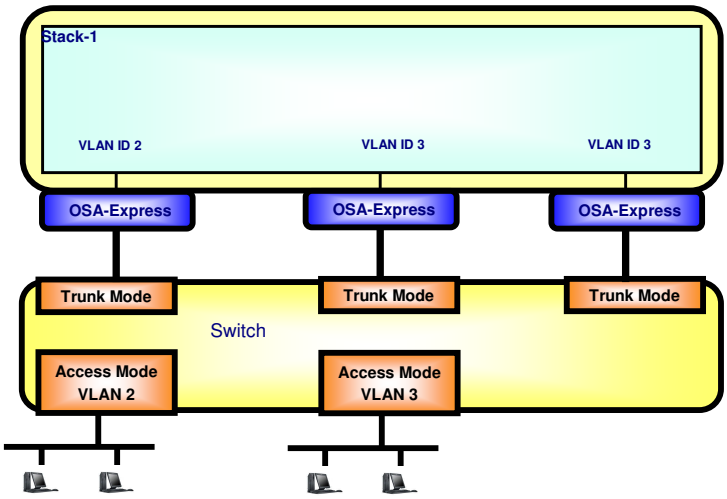
Physical network diagram - Multiple stacks, multiple subnets, multiple OSA ports



Same logical network diagram as on previous page.

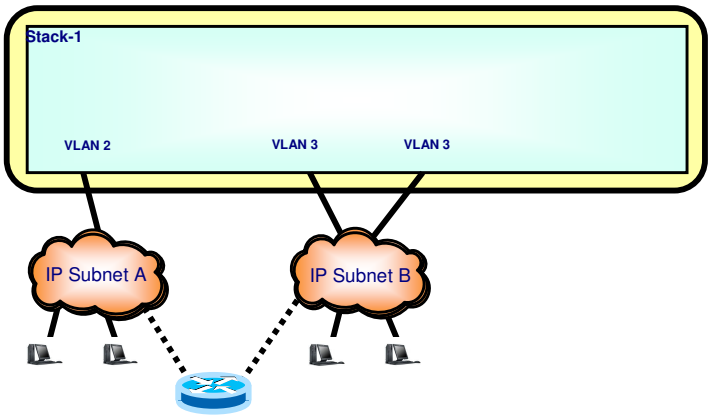
© Copyright International Business Machines Corporation 2004. All rights reserved.

Physical network diagram - Single stack, Multiple subnets, multiple OSA ports



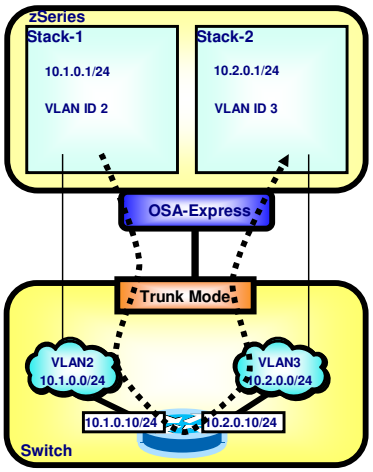
© Copyright International Business Machines Corporation 2004. All rights reserved.

Logical network diagram - Single stack, multiple subnets, multiple OSA ports



© Copyright International Business Machines Corporation 2004. All rights reserved.

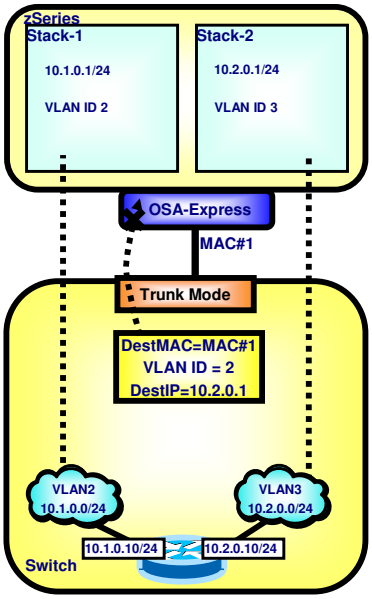
Sharing OSA ports between stacks belonging to different VLAN IDs



- Stack-1 has a global VLAN ID of 2 configured on its IPv4 LINK statement.
- Stack-2 has a global VLAN ID of 3 configured on its IPv4 LINK statement.
- Stack-1 has a routing table entry that points to 10.1.0.10 for forwarding to the 10.2.0.0/24 (VLAN ID 3) subnet.
- Stack-1 sending an IP packet with a destination IP address in the IP header of 10.2.0.1 - tagged with VLAN ID 2 - with next hop IP address 10.1.0.10 (the router interface on VLAN ID 2).
- Which path will this packet take?
  - To the OSA adapter, then on to the router at 10.1.0.10, then routed over the router's other interface to VLAN ID 3, back up to the OSA adapter, and then to Stack-2.
- This is what you want! The reasons for using VLANs is often security - separating different subnets via a router that applies IP filters before forwarding packets.

© Copyright International Business Machines Corporation 2004. All rights reserved.

Inbound processing - unicast frames



- Stack-1 has a global VLAN ID of 2 configured on its IPv4 LINK statement.
- Stack-2 has a global VLAN ID of 3 configured on its IPv4 LINK statement.
- Assume that a unicast frame with an IP packet arrives at MAC#1 with the following characteristics (please note that this could only be the case if a rogue node was attached to the switch using a trunk interface and the node manipulated the frame contents - aka. a hacker!):
  - Destination IP address 10.2.0.1
  - VLAN ID tag of VLAN 2
- What will the OSA adapter do with that IP packet?
  - Discard it since the VLAN ID of the registered IP address (10.2.0.1 - VLAN ID 3) doesn't match the VLAN tag in the frame (VLAN ID 2)

© Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO Capacity Limits

Copyright International Business Machines Corporation 2004. All rights reserved.

QDIO and iQDIO limits increased

- IQD CHPIDs go from 4 to 16 with z990 and z890
- IQD CHPIDs can span CSS (Channel Subsystems) on z990 and z890
- OSD CHPID spanning support coming
- Number of supported stacks (per OSA) increased:
  - OSA multiple control unit support coming (allows more subchannel devices under the same OSA CHPID)
  - 240 devices (current limit with single CU) increased which raises the current limit of 80 stacks to 640 stacks (1920 devices)
    - note - the current 240 limit was per OSA CHPID per CEC - the CEC limit was raised to 480 (GA1) - this change is transparent to software

© Copyright International Business Machines Corporation 2004. All rights reserved.

HiperSockets Broadcast

Copyright International Business Machines Corporation 2004. All rights reserved.

HiperSockets Broadcast

- Problem:
  - Applications which rely on broadcast packets (e.g. RIPv1, DHCP) do not work over HiperSockets interfaces
- Solution:
  - Add support for broadcast traffic over HiperSockets for IPv4 packets
  - Keep HiperSockets functionality as similar as possible to OSA-Express QDIO functionality (Note: TCP/IP added broadcast support for OSA-Express QDIO in V1R4)
- How to enable:
  - New IPBCAST keyword on the IPAQIDIO link statement

```
>>-LINK--link_name--IPAQIDIO--device_name--+---------+--->
                                           '-IPBCAST-'

  .-READSTORAGE-----GLOBAL-.
>-+------------------------+--|
  +-READSTORAGE--+--MAX----+
                 '--AVG----+
                 '--MIN----'
```

- Things to consider:
  - Only need to take action if you want broadcast capability
  - Need IBM eServer zSeries 990 or 890 to get the new function
  - No broadcast support for HiperSockets device used for DYNAMICXCF (IUTIQDIO)

© Copyright International Business Machines Corporation 2004. All rights reserved.

Trademarks, Copyrights, and Disclaimers

e-business

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

IBM	CICS	IMS	MQSeries	Tivoli
IBM(logo)	Cloudscape	Informix	OS/390	WebSphere
e(logo)business	DB2	iSeries	OS/400	xSeries
AIX	DB2 Universal Database	Lotus	pSeries	zSeries

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 registered trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

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

Linux is a registered trademark of Linus Torvalds.

Other company, product and service names may be trademarks or service marks of others.

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 and/or changes in the product(s) and/or program(s) 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 (e.g., 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.

© Copyright International Business Machines Corporation 2005. All rights reserved.

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.

© Copyright International Business Machines Corporation 2004. All rights reserved.