Getting the Most Out of Your OSA Adapter with z/OS Communication Server

Alfred B Christensen – alfredch@us.ibm.com
IBM Raleigh, NC

Session 6869
Tuesday 3-Aug-2010 – 3:00 PM to 4:00 PM

Getting the Most Out of Your OSA Adapter with z/OS Communication Server

Session number:	6869
Date and time:	Tuesday August 3, 2010 - 3:00 PM - 4:00 PM
Location:	Room 109 (Hynes Convention Center)
Program:	Communications Infrastructure
Project:	Communications Server
Track:	Network Support and Management and SNA/IP Integration
Classification:	Technical
Speaker:	Alfred B Christensen, IBM
Abstract:	OSA-Express ports are used for both SNA and IP LAN connectivity to System z operating systems. In this session we will focus on how z/OS Communications Server uses OSA for IPv4 LAN connectivity based on Queued Direct IO (QDIO). The session will discuss both configuration and operational aspects, such as maintenance of the OSA address table (OAT), ARP processing, use of Virtual IP Addresses (VIPA), interface availability, sharing capabilities, VLAN support, Virtual MACs, OSA Direct SNMP, and other management related functions. The session will focus on the software use of OSA and will discuss hardware aspects only where such aspects are of importance to understanding how the hardware and software cooperate to deliver the desired functions.

Trademarks, notices, and disclaimers

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

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

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

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

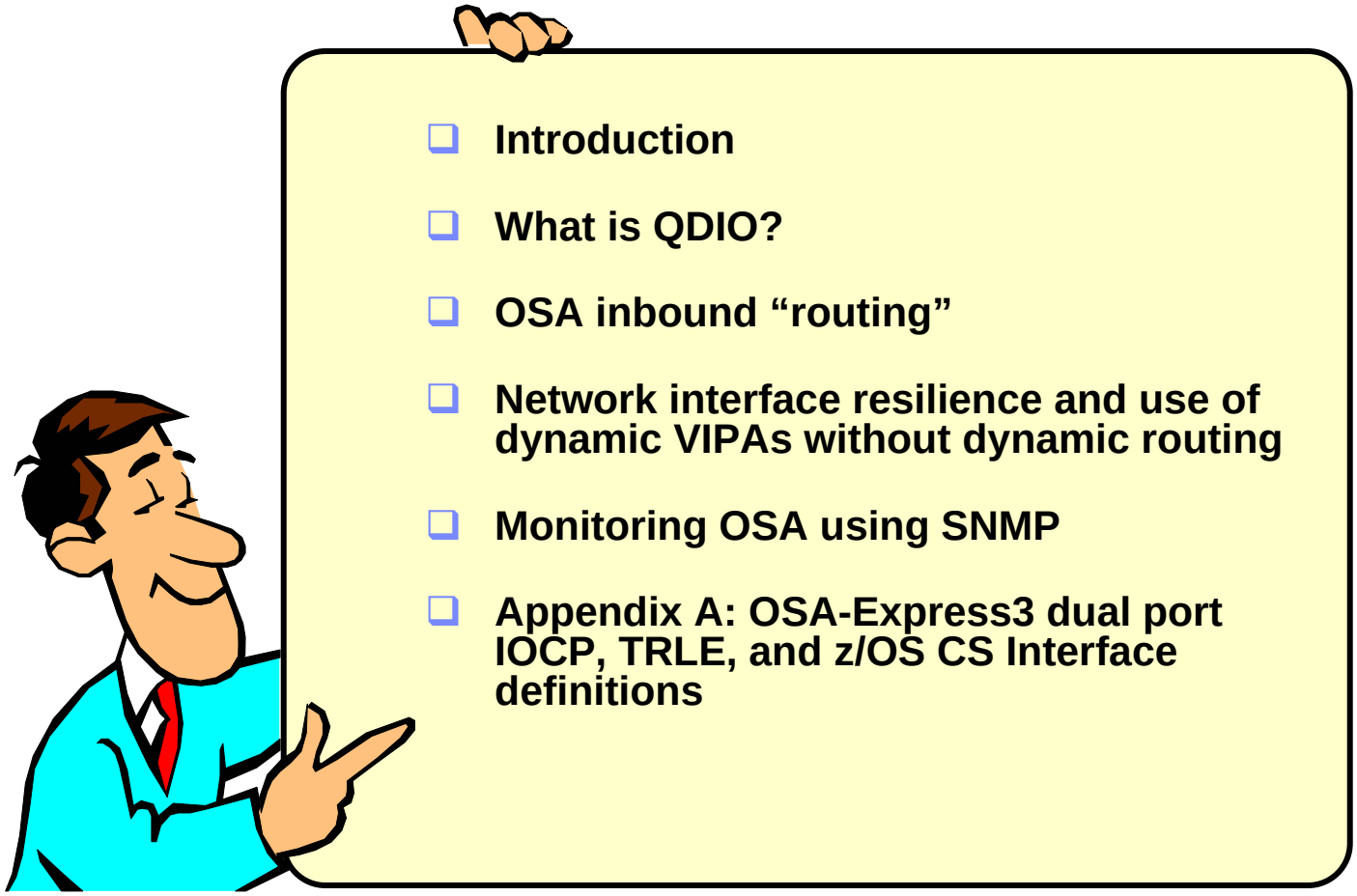
- Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
- Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or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

Notes:

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

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

Agenda



- Introduction
- What is QDIO?
- OSA inbound “routing”
- Network interface resilience and use of dynamic VIPAs without dynamic routing
- Monitoring OSA using SNMP
- Appendix A: OSA-Express3 dual port IOCP, TRLE, and z/OS CS Interface definitions

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.

Getting the most out of your OSA adapter with z/OS Communications Server

Introduction

OSA-Express connectivity overview - EC

Not yet updated with z196!

Feature	Feature Name	Ports	z900	z990	z9 EC	z10 EC	CHPIDs	Connectors
OSA-Express								
2362	155 ATM SM	2	X	RPQ	N / A	N / A	OSD, OSE	SC Duplex
2363	155 ATM MM	2	X	RPQ	N / A	N / A	OSD, OSE	SC Duplex
2364	GbE LX	2	X	C	C	N / A	OSD, L2/L3 **	SC Duplex
2365	GbE SX	2	X	C	C	N / A	OSD, L2/L3 **	SC Duplex
2366	Fast Ethernet	2	X	C	C	N / A	OSD, OSE	RJ-45
2367	Token Ring	2	X	X	N / A	N / A	OSD, OSE	RJ-45
OSA-Express								
1364	GbE LX	2	09/04	X	C	N / A	OSD, L2/L3 **	LC Duplex
1365	GbE SX	2	09/04	X	C	N / A	OSD, L2/L3 **	LC Duplex
1366	1000BASE-T	2	N / A	X	C	N / A	OSC, OSD L2/L3, OSE	RJ-45
OSA-Express2								
3364	GbE LX	2	N / A	01/05	X	X	OSD L2/L3, OSN *	LC Duplex
3365	GbE SX	2	N / A	01/05	X	X	OSD L2/L3, OSN *	LC Duplex
3366	1000BASE-T	2	N / A	05/06	X	X	OSC, OSD L2/L3, OSE, OSN *	RJ-45
3368	10 GbE LR	1	N / A	01/05	X	w/d 30Jun	OSD L2/L3 **	SC Duplex
OSA-Express3								
3362	GbE LX	4	N / A	N / A	N / A	30May08	OSD L2/L3, OSN	LC Duplex
3363	GbE SX	4	N / A	N / A	N / A	30May08	OSD L2/L3, OSN	LC Duplex
3367	1000BASE-T	4	N / A	N / A	N / A	28Oct08	OSC, OSD L2/L3, OSE, OSN *	RJ-45
3370	10 GbE LR	2	N / A	N / A	N / A	30May08	OSD L2/L3	LC Duplex
3371	10 GbE SR	2	N / A	N / A	N / A	28Oct08	OSD L2/L3	LC Duplex

X = Available for ordering
C = Carry forward on an upgrade
* OSN is exclusive to z10 and z9
** L2/L3 = Layer 2/Layer 3 which is applicable to z990 and later servers

OSA-Express connectivity overview - BC

Not yet updated with z196!

Feature	Feature Name	Ports	z800	z890	z9 BC	z10 BC	CHPIDs	Connectors
OSA-Express								
2362	155 ATM SM	2	X	RPQ	N / A	N / A	OSD, OSE	SC Duplex
2363	155 ATM MM	2	X	RPQ	N / A	N / A	OSD, OSE	SC Duplex
2364	GbE LX	2	X	C	C	N / A	OSD, L2/L3 **	SC Duplex
2365	GbE SX	2	X	C	C	N / A	OSD, L2/L3 **	SC Duplex
2366	Fast Ethernet	2	X	C	C	N / A	OSD, OSE	RJ-45
2367	Token Ring	2	X	X	N / A	N / A	OSD, OSE	RJ-45
OSA-Express								
1364	GbE LX	2	09/04	X	C	N / A	OSD, L2/L3 **	LC Duplex
1365	GbE SX	2	09/04	X	C	N / A	OSD, L2/L3 **	LC Duplex
1366	1000BASE-T	2	N / A	X	C	N / A	OSC, OSD L2/L3, OSE	RJ-45
OSA-Express2								
3364	GbE LX	2	N / A	01/05	X	X	OSD L2/L3, OSN *	LC Duplex
3365	GbE SX	2	N / A	01/05	X	X	OSD L2/L3, OSN *	LC Duplex
3366	1000BASE-T	2	N / A	05/06	X	X	OSC, OSD L2/L3, OSE, OSN *	RJ-45
3368	10 GbE LR	1	N / A	01/05	X	C	OSD L2/L3 **	SC Duplex
OSA-Express3								
3362	GbE LX	4	N / A	N / A	N / A	28Oct08	OSD L2/L3, OSN	LC Duplex
3363	GbE SX	4	N / A	N / A	N / A	28Oct08	OSD L2/L3, OSN	LC Duplex
3367	1000BASE-T	4	N / A	N / A	N / A	28Oct08	OSC, OSD L2/L3, OSE, OSN *	RJ-45
3369	2P 1000BASE-T	2	N / A	N / A	N / A	28Oct08	OSC, OSD L2/L3, OSE, OSN *	RJ-45
3370	10 GbE LR	2	N / A	N / A	N / A	28Oct08	OSD L2/L3	LC Duplex
3371	10 GbE SR	2	N / A	N / A	N / A	28Oct08	OSD L2/L3	LC Duplex
3373	2P GbE SX	2	N / A	N / A	N / A	28Oct08	OSD L2/L3, OSN	LC Duplex

X = Available for ordering
C = Carry forward on an upgrade
*** OSN is exclusive to z10 and z9**
**** L2/L3 = Layer 2/Layer 3 which is applicable to z990 and later servers**

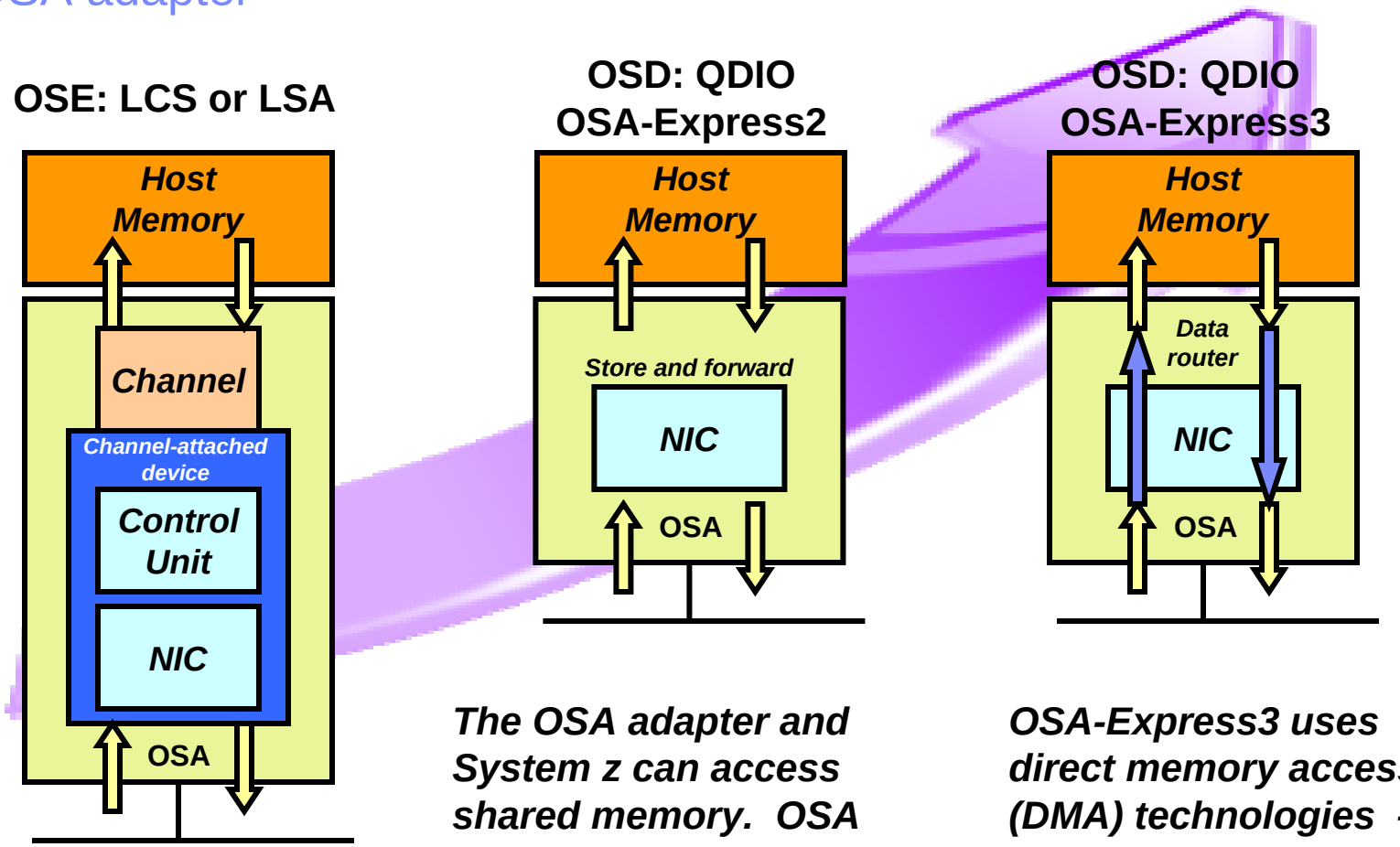
What are the CHPID types used for?

Operating system	CHPID Type:	OSE		OSD		OSN	OSC	OSM	OSX	Comments
	Device driver operation:	LCS	LSA	QDIO L3	QDIO L2	CDLC	(CDLC)	QDIO L3	QDIO L3	
z/OS	VTAM SNA LLC2 LAN		✓							
	VTAM SNA CDLC					✓				
	IPv4	✓		✓					✓	
	IPv6			✓				✓	✓	
	z/OS NIP/MCS Console						✓			
	z/OS non-SNA DFT VTAM						✓			VTAM Local 3270 major node
z/VM	VTAM SNA LLC2 LAN		✓							
	VTAM SNA CDLC					✓				
	IPv4	✓		✓					✓	
	IPv6			✓				✓	✓	z/VM 5.1
	z/VM Virtual Switch			✓	✓					z/VM 5.1 for layer 2 support
	z/VM IPL Console						✓			
	z/VM 3270 terminal						✓			
Linux on System z	CSL SNA LLC2 LAN	✓			✓					CSL 6.2.1
	CCL SNA LLC2 LAN	✓			✓					CCL 1.2, Linux 2.6
	CCL SNA CDLC					✓				CCL 1.2, Linux 2.6
	IPv4	✓		✓	✓				✓	
	IPv6			✓	✓			✓	✓	
z/VSE	VTAM SNA LLC2 LAN		✓							
	VTAM SNA CDLC					✓				
	IPv4	✓		✓						
	z/VSE Console						✓			
	z/VSE non-SNA DFT VTAM						✓			Assume so, no doc found
z/TPF	IPv4			✓						
	SNA CDLC					✓				
	z/TPF Console						✓			

Getting the most out of your OSA adapter with z/OS Communications Server

What is QDIO?

Queued Direct IO (QDIO) – how System z communicates with the LAN via an OSA adapter



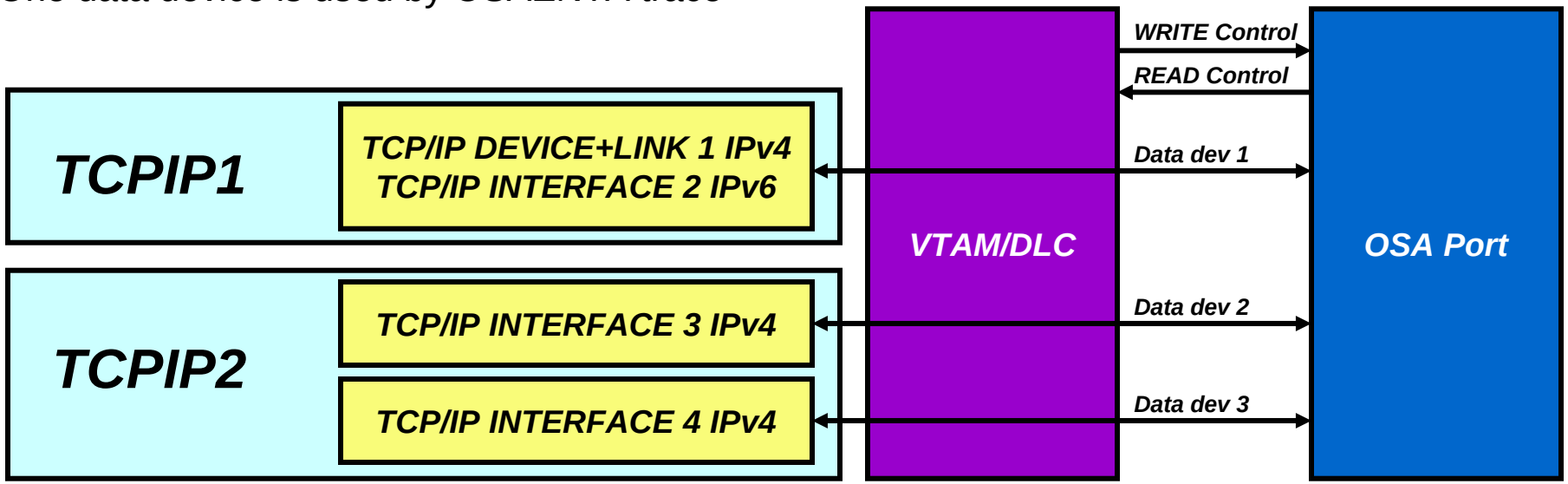
The OSA adapter and System z can access shared memory. OSA interrupts System z using PCI, while System z "interrupts" OSA via a SIGA instruction.

OSA-Express3 uses direct memory access (DMA) technologies – allowing data to flow through OSA without slow store and forward processing.

A little more on QDIO and devices

- Control data between z/OS CS and the OSA port is exchanged over one read device and one write device (READ even address, WRITE odd address)
- The READ and WRITE device pair is shared among all the IP interface definitions in an LPAR that use the same OSA port
- Data is exchanged over data devices
- One data device is used per:
  - A set of exactly one IPv4 and one IPv6 interface
    - IPv4: DEVICE+LINK, IPv6: INTERFACE stmt.
  - One IPv4 interface
  - One IPv6 interface
- One data device is used by OSAENTA trace

```
LNCTL=MPC,
READ=(2FD6),
WRITE=(2FD7),
MPCLEVEL=QDIO,
DATAPATH=(2FD8,2FD9,2FDA,2FDB),
PORTNAME=(O3ETHB1P),
PORTNUM=(1)
```



QDIO and OSA are about much more than plain throughput

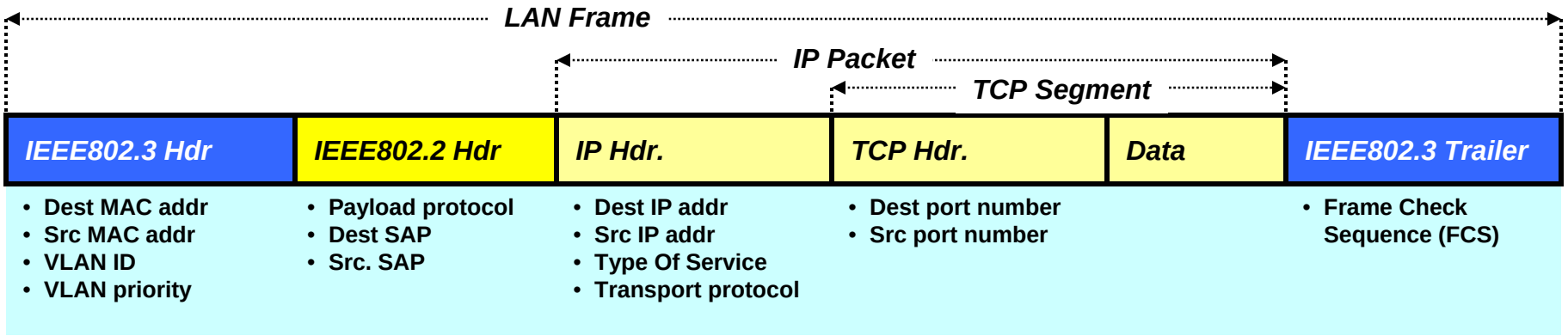
- Home IP address registration (OSA Address Table – also known as OAT)
- Default OSA router registration (PRI/SEC ROUTER)
- Virtual MAC address registration (VMAC)
- VLAN support incl. VLAN priority and dynamic VLAN registration via GVRP
- ARP offload
- Checksum offload
- TCP segmentation offload
- MAC and LLC header construction offload
- Broadcast and multicast support
- OSA Direct SNMP interface
- Network Traffic Analyzer (built-in “sniffer” like LAN tracing)
- OSA Express3 Optimized Latency Mode (OLM)
- QDIO Layer-2 mode (Network protocol layer agnostic)
- Outbound priority queues (four priority queues)
- LPAR-to-LPAR communication without sending data over the LAN
- …..

Getting the most out of your OSA adapter with z/OS Communications Server

OSA inbound “routing”

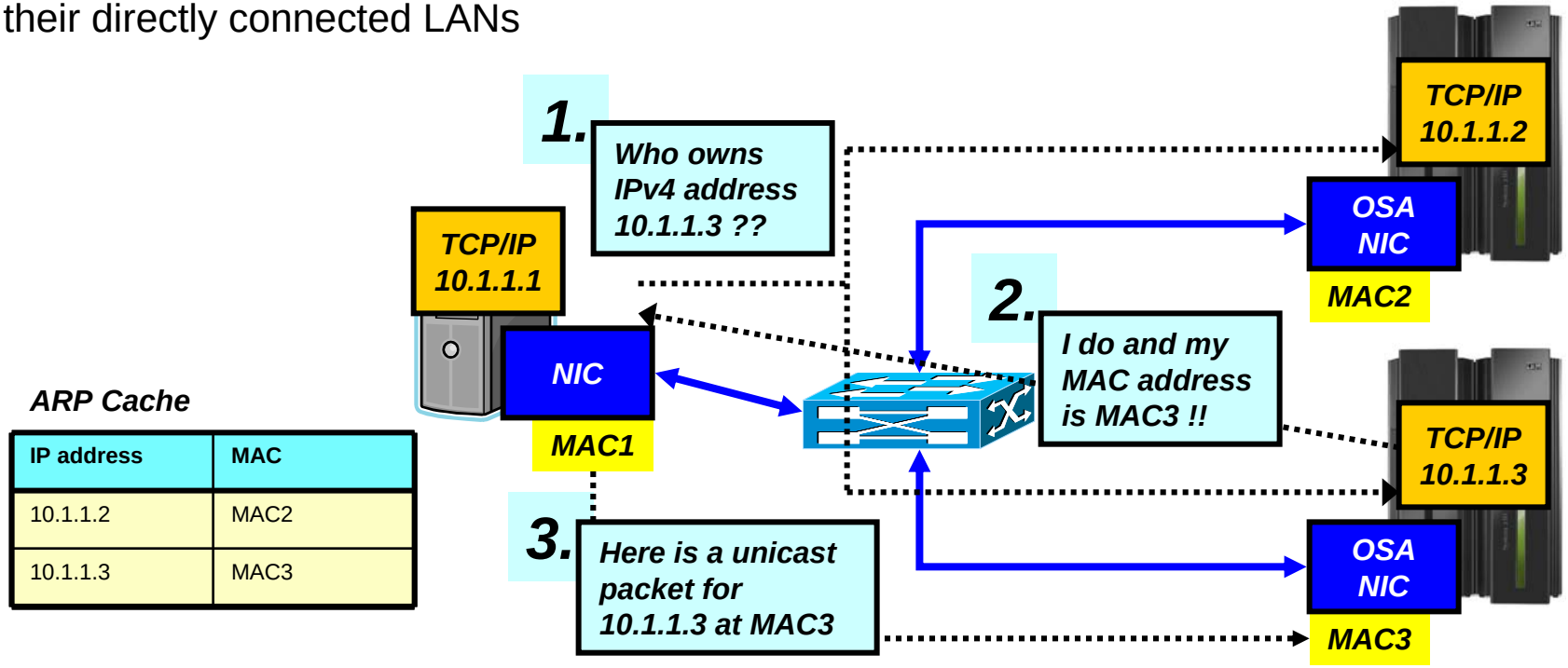
Some basic LAN technology overview

- The LAN infrastructure transports “Frames” between Network Interface Cards (NICs) that are attached to the LAN media (Copper or fiber optic)
- Each NIC has a hardware address
  - A Media Access Control (MAC) address
    - Burned in (world-wide unique by vendors) or alternatively locally administered
- Every frame comes from a MAC and goes to a MAC
  - There are special MAC values for broadcast and multicast frames
- Every frame belongs to the physical LAN or to one of multiple Virtual LANs (VLAN) on the physical LAN
  - A VLAN ID is in the frame header if VLAN technologies are in use
- A frame carries a payload of a specified protocol type, such as ARP, IPv4, IPv6, SNA LLC, etc.

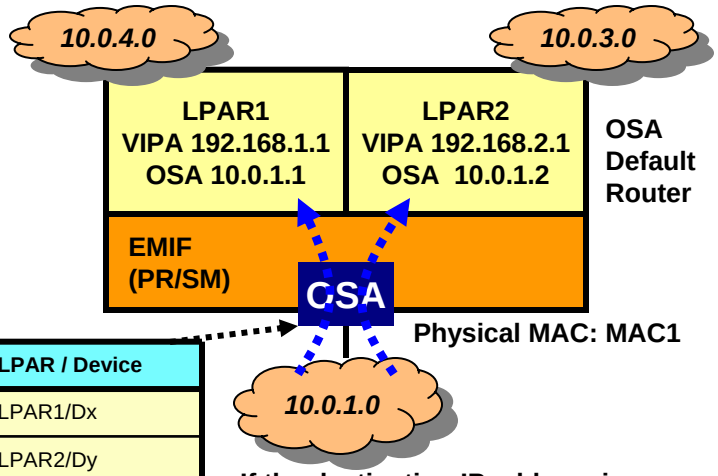


Correlation of IPv4 addresses and MAC addresses on a LAN – Address Resolution Protocol (ARP)

- An IPv4 node uses the ARP protocol to discover the MAC address of another IPv4 address that belongs to the same IPv4 subnet as it does itself.
- ARP requests are broadcasted to all NICs on the LAN
- The one NIC that has a TCP/IP stack with the requested IPv4 address responds directly back to the IPv4 node that sent out the broadcast
- Each IPv4 node maintains a cache of IPv4 addresses and associated MAC addresses on their directly connected LANs



So how does this work in a System z environment with shared (virtualized) OSA adapters?



OSA Address Table (OAT)

IP address	LPAR / Device
192.168.1.1	LPAR1/Dx
192.168.2.1	LPAR2/Dy
10.0.1.1	LPAR1/Dx
10.0.1.2	LPAR2/Dy
Primary	LPAR2/Dy
Secondary	LPAR1/Dx

If the destination IP address in a frame that arrives at the OSA adapter doesn't belong to any of the LPARs that share the adapter, that IP packet is given to the LPAR that acts as the Default OSA router. There are separate IPv4 and IPv6 default OSA router specifications.

LAN frames

Dest_MAC	Dest_IPv4	Who gets it?	Why?
MAC1	10.0.1.1	LPAR1	Registered
MAC1	192.168.1.1	LPAR1	Registered
MAC1	10.0.1.2	LPAR2	Registered
MAC1	192.168.2.1	LPAR2	Registered
MAC1	10.0.3.5	LPAR2	Default router
MAC1	10.0.4.6	LPAR2	Default router

- An OSA NIC has a physical MAC address, just like all NICs
- An OSA NIC is often used by multiple LPARs and TCP/IP stacks
  - The NIC is virtualized so it functions as the NIC for multiple TCP/IP stacks, each with their own IP address
- If someone ARPs for 10.0.1.1 in LPAR1, OSA will return MAC1
- If someone ARPs for 10.0.1.2 in LPAR2, OSA will also return MAC1
- So what does OSA then do when a unicast frame arrives with a destination MAC address of MAC1?
  - It peeks into the IP header inside the frame, and consults a table known as the OSA Address Table (OAT) to see which LPAR the IP address inside the frame belongs to

Setting a few facts straight about OSA and “routing”

- **OSA is not a full-function IP router.**
  - OSA can analyze a destination IP address in a LAN frame to decide which LPAR an incoming IP packet belongs to.
  - OSA soes not participate in dynamic routing updates.
  - OSA does not act as a general IP router
    - Depends on the System z TCP/IP stacks to do that and handle all IP routing issues.

- **Originally, the OSA NIC has a single physical MAC address that is shared among all the LPARs that share the OSA port.**
  - All IP packets to all LPARs can be destined for one and the same MAC address
  - In that case, OSA selects stack based on destination IP address and the OSA Address Table

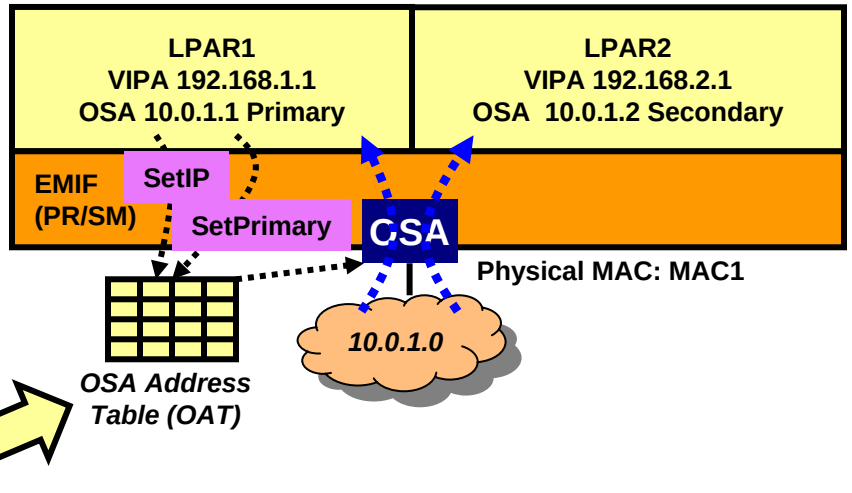
- **If z/OS acts as an IP router to IP networks behind z/OS, the destination address may be any IP address on those networks behind z/OS.**
  - LPAR designated as default router will receive such packets
  - Can become extremely cumbersome to set up if LPARs that share an OSA port are connected to different back-end networks

- **When a port is defined as an OSE/LCS port, the content of the OAT must be maintained *manually* through OSA/SF interaction.**

- **When a port is defined as a OSD/QDIO port, the content of the OAT is maintained *automatically* by the sharing TCP/IP stacks.**

QDIO layer-3: less administration, more dynamics – but OAT remains an important element of OSA inbound routing

QDIO Layer 3 - the OSA adapter is IP-aware and offloads certain functions from the TCP/IP stacks – check-summing, ARP processing, TCP segmentation, etc.



OAT is maintained dynamically by TCP/IP

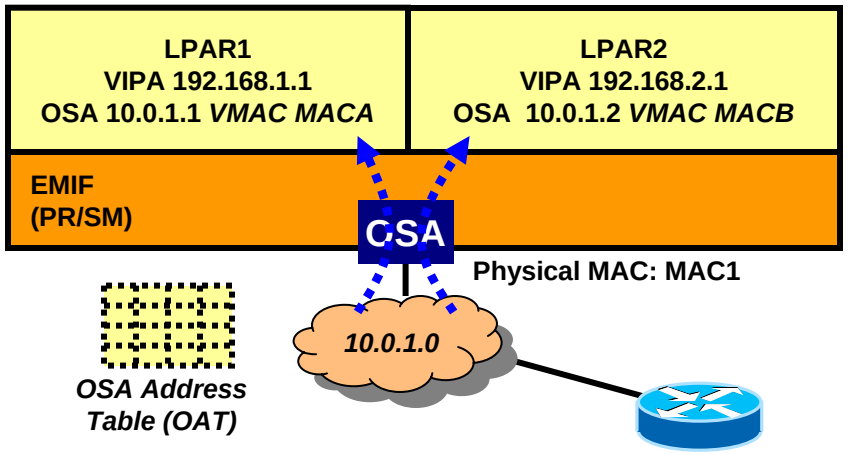
IP address	LPAR / Device
192.168.1.1	LPAR1/Dx
192.168.2.1	LPAR2/Dy
10.0.1.1	LPAR1/Dx
10.0.1.2	LPAR2/Dy
Primary	LPAR1/Dx
Secondary	LPAR2/Dy

OAT Updates:

- **LCS: operator intervention via OSA/SF commands**
- **QDIO: dynamic by software via QDIO control channel**

- QDIO device definitions in the TCP/IP stacks are used to dynamically establish the stack as the OAT default router, secondary router, or non-router.
- Whenever a QDIO device is activated or the TCP/IP home list is modified (through OBEYFILE command processing or through dynamic changes, such as dynamic VIPA takeover), the TCP/IP stack updates the OAT configuration dynamically with the HOME list IP addresses of the stack.
- The OAT includes all (non-LOOPBACK) HOME IP addresses of all the stacks that share the OSA adapter.
- The fact that the OSA microcode is IP address-aware (as it is in this scenario) is the reason for referring to this as QDIO layer 3 processing (layer 3 is generally the networking layer in an OSI model - the IP networking layer when using TCP/IP)

Wouldn't it be so much easier with a MAC address per z/OS TCP/IP network interface?



YES !!

Router's ARP cache

IP Address	MAC Address
10.0.1.1	MACA
10.0.1.2	MACB

The whole mess with PRIROUTER and SECROUTER is gone !!

It also removes issues with external load balancers that use MAC-level forwarding. It makes a system z LPAR look like a "normal" TCP/IP node on a LAN.

- A packet for 192.168.1.1 arrives at the router from the downstream network
  - Router determines the destination is not directly attached to the router
  - Router looks into its routing table and determines next-hop IP address for this destination is 10.0.1.1, which is on a network that is directly attached to the router
  - Router looks into its ARP cache and determines the associated MAC address is MACA
  - Router forwards the IP packet in a LAN frame to MACA
- Frame arrives in OSA
  - OSA determines which LPAR it belongs to based on the virtual MAC address
  - The OAT may optionally still play a role in inbound routing decisions by the OSA adapter:
    - If the destination address (192.168.1.1) were not in the home list of this LPAR (and hence not in the OAT for this LPAR), should OSA still send it up to the LPAR or not?
    - You decide via a configuration option
      - ROUTEALL: send all packets with my VMAC to me
      - ROUTELCL: send only packets to me if they are in my HOME list
  - The OAT remains in use for other functions, such as ARP ownership, etc.

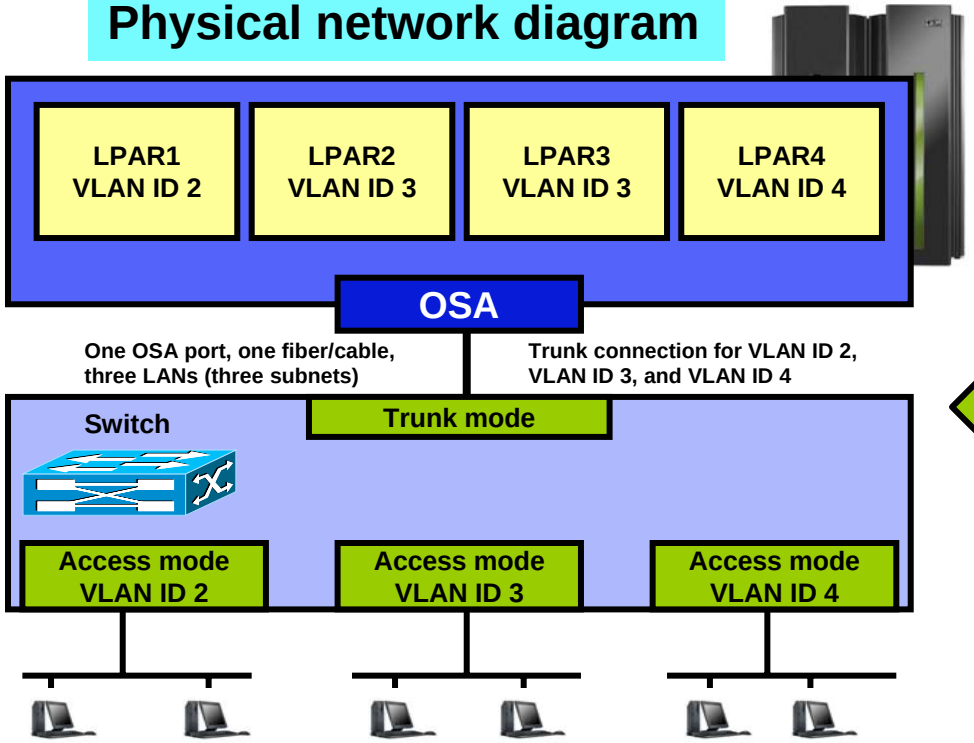
OSA-Express virtual MAC while operating in QDIO layer-3 mode

- **OSA MAC sharing problems do not exist if each stack has its own MAC**
  - "virtual" MAC
  - To the network, each stack appears to have a dedicated OSA port
- **MAC address selection**
  - Coded in the TCP/IP profile
  - Generated and assigned by the OSA adapter
- **All IP addresses for a stack are advertised with the virtual MAC**
  - by OSA using ARP for IPv4
  - by the stack using ND for IPv6
- **All external routers now forward frames to the virtual MAC**
  - OSA will “route” to an LPAR/Stack by virtual MAC instead of IP address
  - All stacks can be "routing" stacks instead of 1 PRIROUTER stack
- **Simplifies configuration greatly**
  - No PRIROUTER/SECROUTER!
- **Supported on System z9, z10, and z196**
- **Prior to z/OS V1R10, each stack may define one VLAN per protocol (IPv4 or IPv6) for each OSA port**
  - One VMAC for the LINK statement
  - One VMAC for the INTERFACE statement
- **z/OS V1R10 allows each stack to define up to 8 VLANS per protocol, each with its own VMAC for each OSA port**
  - A total of 8 IPv4 and 8 IPv6 logical network interfaces per stack

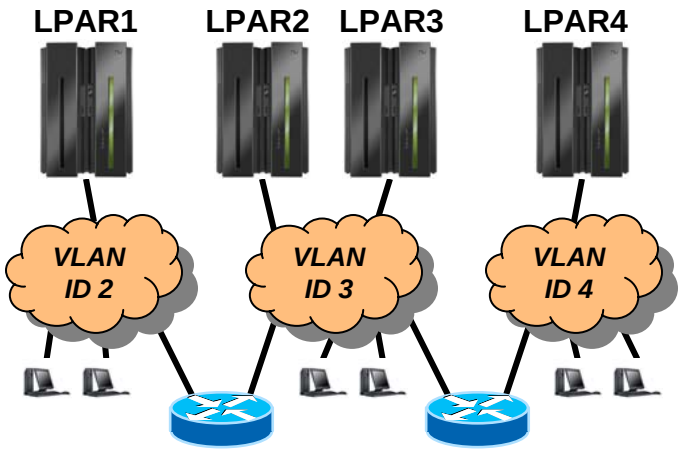
I see no general reasons to not use VMACs. Use them!!!

What is a Virtual LAN (VLAN)?

Physical network diagram



Logical network diagram



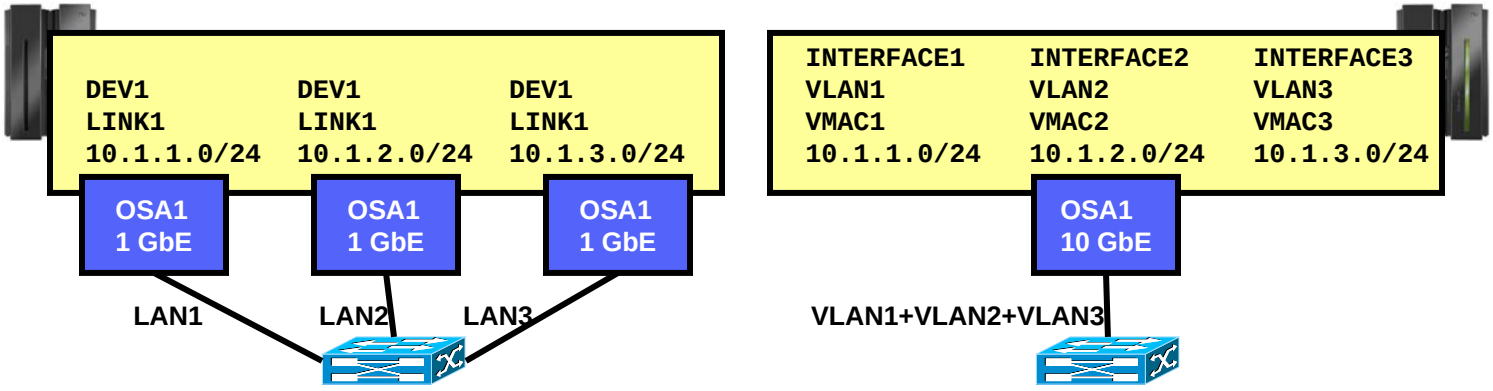
- Depending on switch configuration, the switch may interconnect the VLANs using a layer-3 IP router function.
- The subnets may belong to different routing domains or OSPF areas:
  - Test, production, demo
- The subnets may belong to different security zones:
  - Intranet, DMZ

- **Each frame on the trunk mode connection carries a VLAN ID in the IEEE802.3 header that allows the network equipment to clearly identify which virtual LAN each frame belongs to.**
- **On an access mode connection, the switch will transport frames belonging to the configured VLAN ID for that access mode connection only.**

Note: It is generally a good idea on z/OS to use VLAN in combination with VMACs to avoid inherent complexity issues with PRIROUTER/SECROUTER support in a multi VLAN environment.

Multiple network interfaces (VLANs) per OSA port per stack per IP protocol version

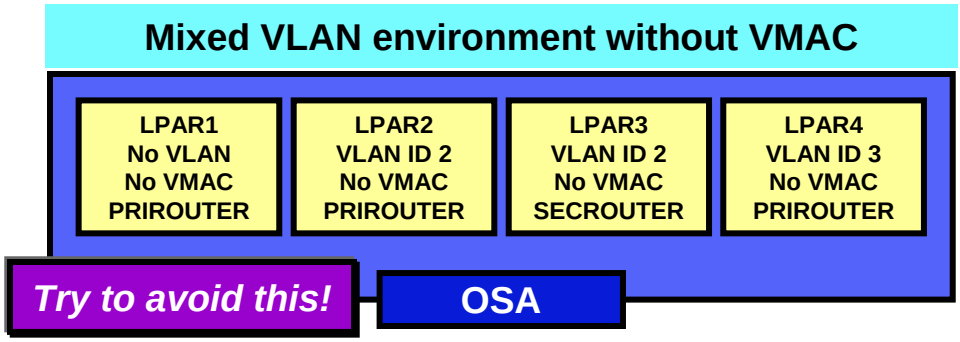
- As installations consolidate multiple OSA Gigabit Ethernet ports to a smaller number of 10 Gigabit Ethernet ports this limitation has become too restrictive:
  - Not possible to retain existing network interface and IP subnet topology
  - Consolidating multiple LANs to one LAN requires IP renumbering
- z/OS V1R10 added support for multiple VLANs per IP protocol per OSA port:
  - Each VLAN on the same OSA port must use unique, non-overlapping IP subnets or prefixes
    - Will be enforced by the TCP/IP stack
  - Each VLAN must be defined using theIPv4-enabled version of the INTERFACE configuration statement
    - IPv4 INTERFACE statement only supports QDIO interfaces
  - Each VLAN must use layer-3 virtual MAC addresses with ROUTEALL, and each VLAN must have a unique MAC address



Consolidate multiple low-capacity LAN interfaces to fewer high-capacity LAN interfaces without network topology impact.

A few more words on VLANs with z/OS CS

- z/OS Communications Server supports one VLAN ID per network interface
  - This is known as a global VLAN ID
    - Remember: z/OS CS now supports up to eight interfaces per OSA port – each with its own VLAN ID
  - OSA performs inbound routing based on VLAN IDs in the incoming frames and only sends frames to z/OS with the global VLAN ID z/OS has registered
- Linux on System z supports multiple VLAN IDs per interface
  - OSA sends multiple VLAN IDs up to Linux and Linux then de-multiplexes the different virtual LANs
- Some interfaces that share an OSA port may select to not specify a VLAN ID, while others do specify a VLAN ID (a mixed VLAN environment)
  - **Warning**: be very careful - mixed VLAN environments are not recommended due to complexities with OSA default router selection and other functional issues. Use VLAN IDs on all TCP/IP interfaces that share an OSA port - or not at all.

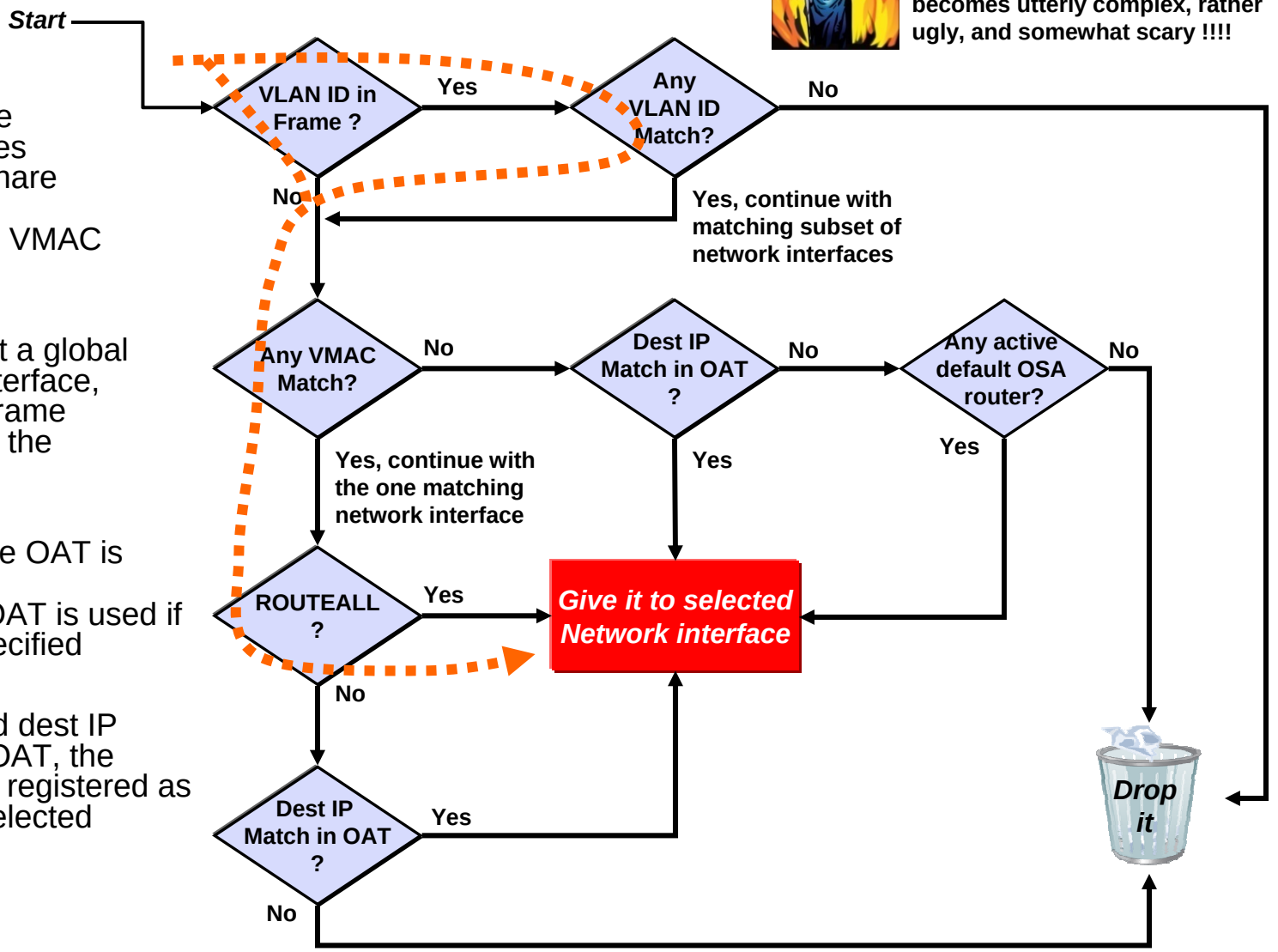


- **LPAR1 acts as PriRouter for:**
  - **Untagged frames**
  - **Frames tagged with a null VLAN ID**
  - **Frames tagged with an unregistered (anything but VLAN 2 or 3) VLAN ID**
  - **Frames tagged with a registered VLAN ID, an unknown destination IP address but no VLAN Pri/SecRrouter**
- **LPAR2 acts as PriRouter for frames tagged with VLAN 2 while LPAR3 acts as SecRouter for that same VLAN.**
- **LPAR4 acts as PriRouter for frames tagged with VLAN 3**

OSA inbound routing as of early 2010

This is correct if all interfaces that share the OSA port use VLAN IDs or none of them do. If some do and others do not, this becomes utterly complex, rather ugly, and somewhat scary !!!!

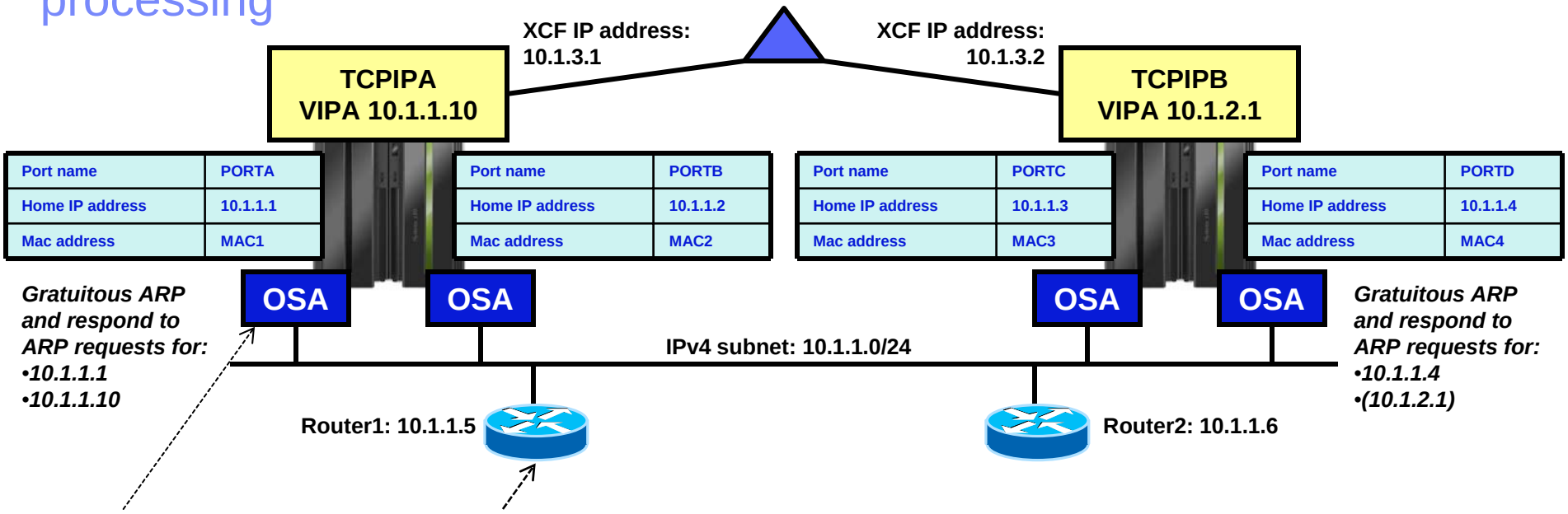
- **Dest MAC**
  - VMACs are unique across all interfaces and VLANs that share an OSA port
  - z/OS CS registers VMAC with the OSA port
- **VLAN ID**
  - If z/OS CS has set a global VLAN ID for an interface, OSA verifies the frame VLAN ID matches the global VLAN ID
- **Dest IP address**
  - Without VMAC, the OAT is always used
  - With VMAC, the OAT is used if ROUTELCL is specified
- **Default OSA router**
  - If OAT is used and dest IP address is not in OAT, the interface currently registered as default router is selected



Getting the most out of your OSA adapter with z/OS Communications Server

Network interface resilience and use of dynamic VIPAs without dynamic routing

QDIO Layer 3 basics with respect to VIPA addresses and ARP processing



OSA PORTA's OAT

IP Address	ARP Owner
10.1.1.1	Yes
10.1.1.10	Yes
10.1.1.2	No
10.1.3.1	No

Router1's ARP cache

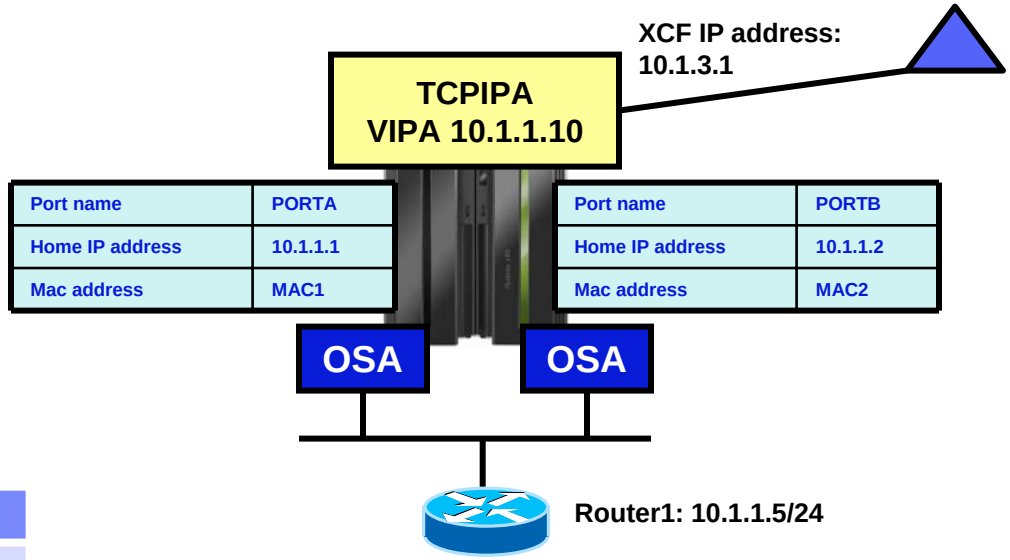
IP Address	MAC Address
10.1.1.1	MAC1
10.1.1.2	MAC2
10.1.1.3	MAC3
10.1.1.4	MAC4
10.1.1.10	MAC1

- With VMAC and ROUTEALL only addresses for which OSA has to perform ARP are registered in the OAT
- In all other cases, all HOME IP addresses will be registered in the OAT and the OAT content will be changed as the HOME lists change due to (dynamic) movement of IP addresses.
- For QDIO interfaces that are defined with DEVICE and LINK statement, OSA will do gratuitous ARP for the OSA interface IP address and all VIPA addresses.
- For QDIO interfaces that are defined with the INTERFACE statement (z/OS V1R10+), OSA will do gratuitous ARP for the OSA interface IP address and for VIPA addresses that belong to the same subnet as the OSA interface.

Recommendation: Define IPv4 QDIO interfaces with the INTERFACE statement if you are at z/OS V1R10 or later !

Network connectivity resilience without dynamic routing

z/OS TCP/IP support interface recovery if multiple network interfaces to the same subnet exist. In this example, both OSA PORTA and PORTB are connected to the 10.1.1.0/24 subnet.



When PORTA fails, PORTB is given ARP ownership of the addresses PORTA previously had. PORTB sends gratuitous ARPs to enable downstream routers to update their ARP cache.
Note that downstream routers normally will ignore gratuitous ARPs for IP addresses that do not belong to the subnet on that physical network (in this example the 10.1.1.0/24 subnet)

OSA PORTA fails

Router1's ARP cache

IP Address	MAC Address
10.1.1.1	MAC1
10.1.1.2	MAC2
10.1.1.10	MAC1

Router1's ARP cache

IP Address	MAC Address
10.1.1.1	MAC2
10.1.1.2	MAC2
10.1.1.10	MAC2

OSA PORTA's OAT

IP Address	ARP Owner
10.1.1.1	Yes
10.1.1.10	Yes
10.1.1.2	No
10.1.3.1	No

OSA PORTB's OAT

IP Address	ARP Owner
10.1.1.1	No
10.1.1.10	No
10.1.1.2	Yes
10.1.3.1	No

OSA PORTA's OAT

IP Address	ARP Owner
10.1.1.1	Yes
10.1.1.10	Yes
10.1.1.2	No
10.1.3.1	No

OSA PORTB's OAT

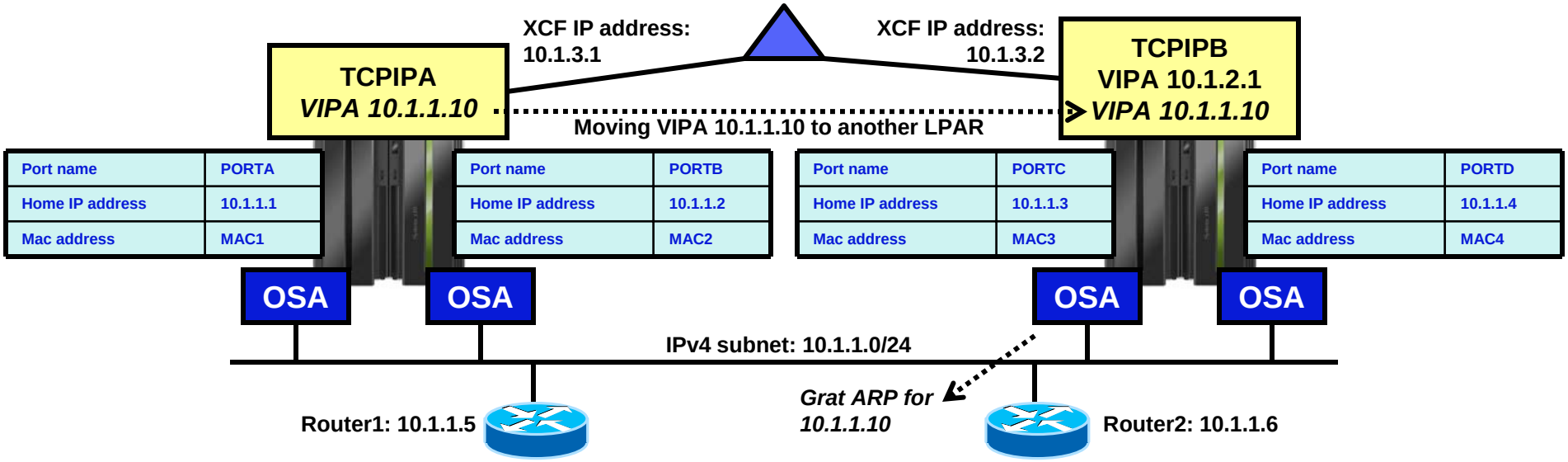
IP Address	ARP Owner
10.1.1.1	Yes
10.1.1.10	Yes
10.1.1.2	Yes
10.1.3.1	No

How do you know which interfaces are on the same subnet and which interface currently handles ARP for your VIPA addresses?

- Messages are issued when an interface takes over ARP responsibility
  - EZD0040I INTERFACE OSAQDIO2 HAS TAKEN OVER ARP RESPONSIBILITY FOR INACTIVE INTERFACE OSAQDIO1
- Messages are issued whenever a previously taken over link or interface recovers and takes back the ARP responsibility.
  - EZD0041I INTERFACE OSAQDIO1 HAS TAKEN BACK ARP RESPONSIBILITY FROM INTERFACE OSAQDIO2
- Use Netstat DEvlinks/-d report to tracks the state of takeover:
  - Displays ARP/ND information.
  - LAN group membership is determined dynamically per interface during interface initialization
  - LAN group numbers are determined dynamically, they are not configured

```
.........
 IPv4 LAN Group Summary
 LanGroup: 00003
   LnkName            LnkStatus          ArpOwner           VipaOwner
   -------            ---------          --------           ---------
   LOSAQDIO4          Active             LOSAQDIO4          Yes
   LOSAQDIO7          Active             LOSAQDIO7          No
 LanGroup: 00004
   LnkName            LnkStatus          ArpOwner           VipaOwner
   -------            ---------          --------           ---------
   LOSAQDIO2          Active             LOSAQDIO2          Yes
```

Dynamic VIPA movement without dynamic routing



When 10.1.1.10 is added to OSA PORTC's OAT with ARP ownership, PORTC sends out a gratuitous ARP that forces nodes on the LAN with that IP address in their ARP cache to update their ARP cache entry with the new MAC address.

OSA PORTA's OAT

IP Address	ARP Owner
10.1.1.1	Yes
10.1.1.2	No
10.1.3.1	No

Router1's ARP cache

IP Address	MAC Address
10.1.1.1	MAC1
10.1.1.2	MAC2
10.1.1.10	***MAC1***
10.1.1.3	MAC3
10.1.1.4	MAC4

OSA PORTC's OAT

IP Address	ARP Owner
10.1.1.3	Yes
10.1.1.10	***Yes***
10.1.1.4	No
10.1.3.2	No
10.1.2.1	No

Router1's ARP cache

IP Address	MAC Address
10.1.1.1	MAC1
10.1.1.2	MAC2
10.1.1.10	***MAC3***
10.1.1.3	MAC3
10.1.1.4	MAC4

Some of the rules for availability and DVIPA movement without dynamic routing

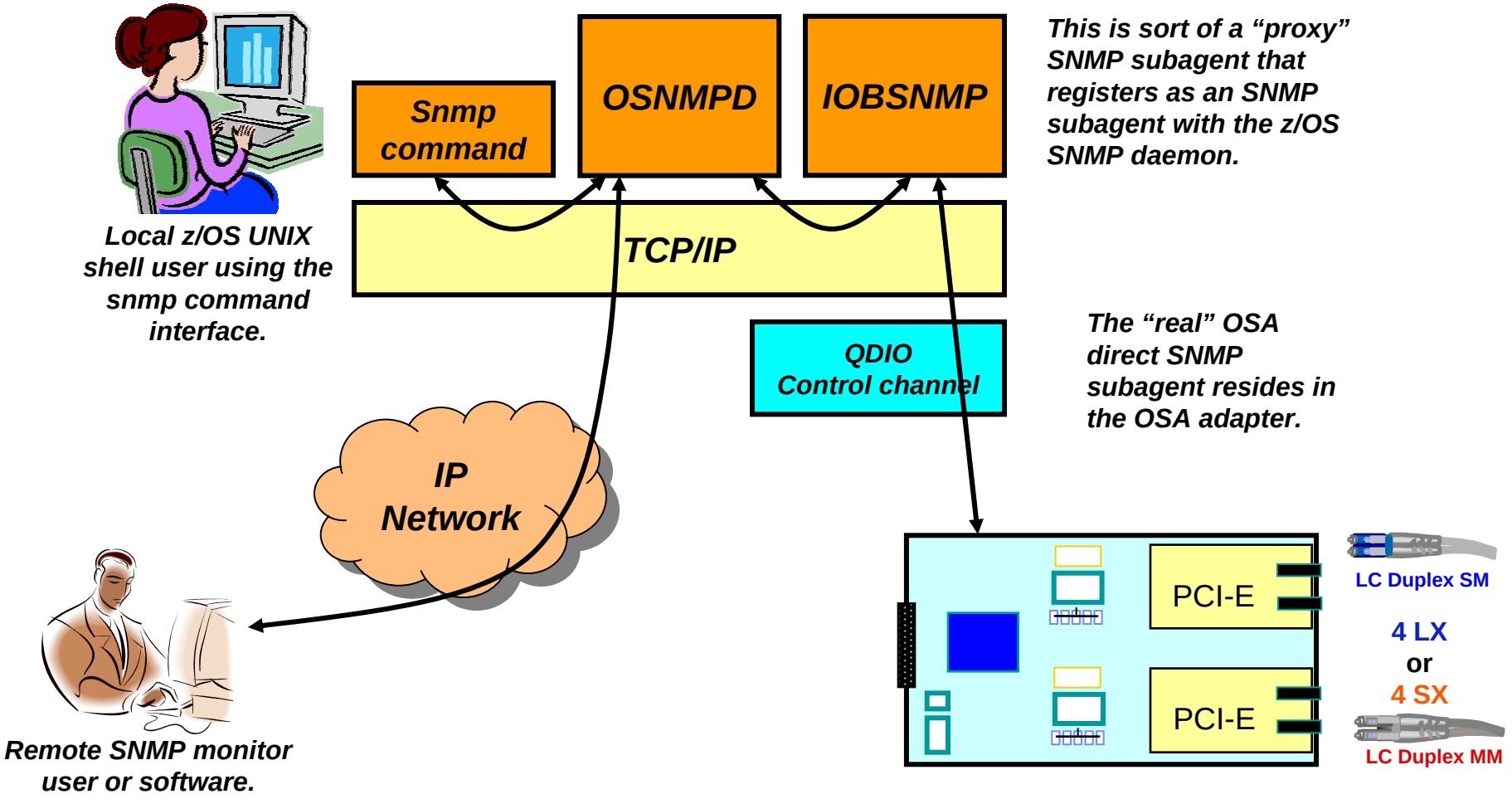
- z/OS VIPA addresses in a flat network configuration without dynamic routing must be allocated out of the same subnet as the directly attached network - in this example, the 10.1.1.0/24 subnet.
- All LPARS in the Sysplex must be attached to one and the same IP subnet via OSA ports.
- Network interfaces belonging to other IP subnets cannot be used for automatic re-routing around failed OSA ports.
  - That includes MPC links, XCF links, or other OSA-attached subnets
- Overall physical availability of the network to which the OSA ports are attached becomes of outmost importance and must generally be based on what is known as Layer2-and-below availability functions in the switches and the physical links (cables).
  - Redundant switches with trunk links
  - Redundant OSA adapters in each LPAR
  - OSA port cabling to multiple switches

These limitations are the reasons why we continue to generally recommend using a dynamic routing infrastructure where z/OS is able to inform downstream routers about interface availability and DVIPA location via dynamic routing protocol updates.

Getting the most out of your OSA adapter with z/OS Communications Server

Monitoring OSA using SNMP

OSA Direct SNMP component overview



Local z/OS UNIX shell user using the snmp command interface.

This is sort of a "proxy" SNMP subagent that registers as an SNMP subagent with the z/OS SNMP daemon.

The "real" OSA direct SNMP subagent resides in the OSA adapter.

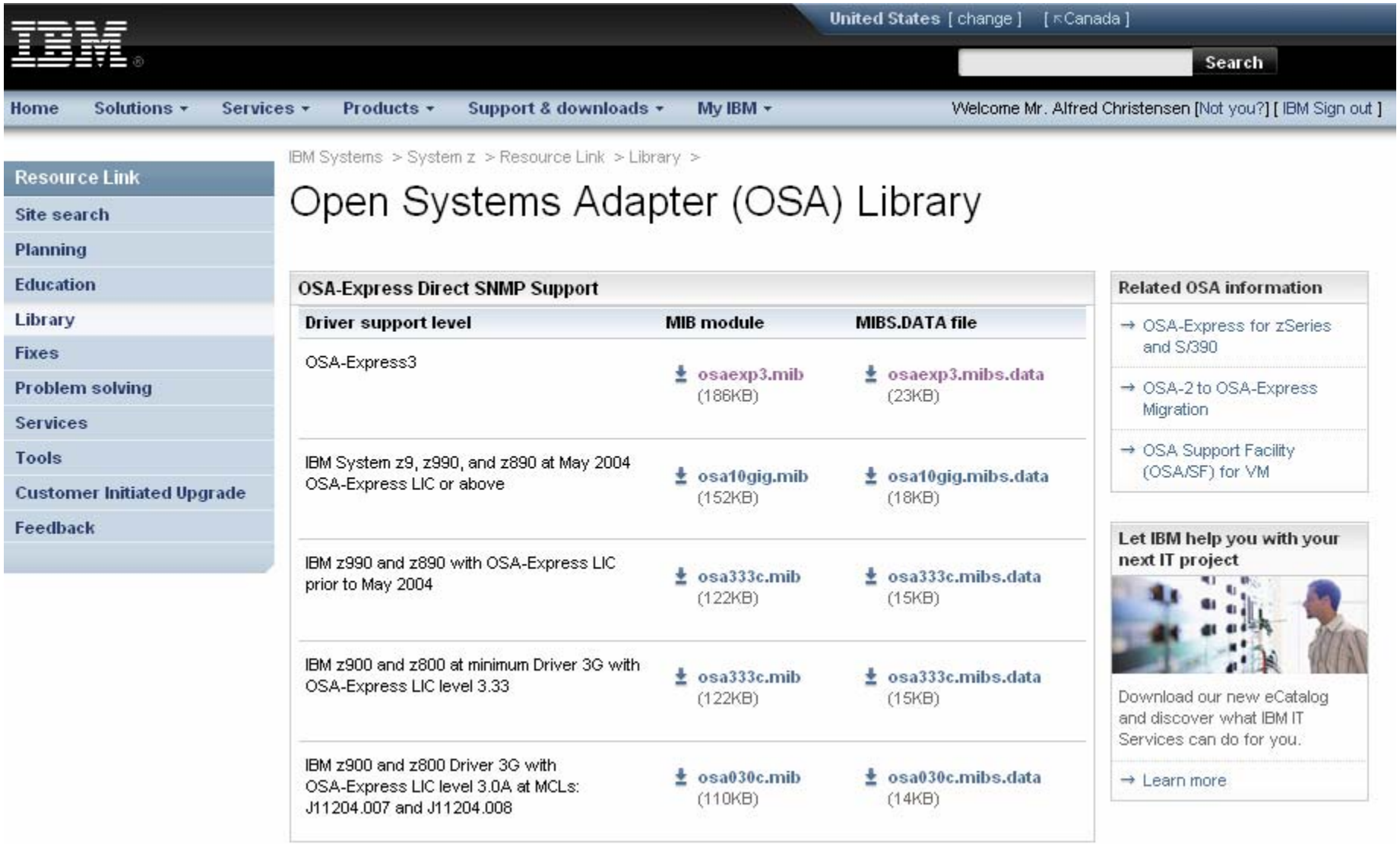
Remote SNMP monitor user or software.

OSA Direct SNMP MIB definitions

- Management Information Base (MIB) variables for the IBM OSA-Express Direct SNMP solution is described:
  - MIB ASN.1 description (and documentation): osaexp3.mib
  - MIB Object ID (OID) to MIB variable name mapping: osaexp3.mibs.data

- These files need to be downloaded from the Web:
  - Point your browser to IBM Resourcelink and log in: https://www.ibm.com/servers/resourcelink/svc03100.nsf?Opendatabase
    - a registration is required to use IBM resourcelink
  - Select "Library“ on the left side
  - Then look to the right side under "Library short cuts" and select “Open System Adapter (OSA) library"
  - At this point you can select the "OSA-Express Direct SNMP MIB module”
  - You might consider signing up to monitor this site to receive notification when changes occur
  - The osaexp3.mibs.data file need to be installed into /etc/mibs.data or appended to an existing /etc/mibs.data file in order to use the z/OS snmp command with names instead of the (cryptic) OID

SNMP files to download from IBM Resourcelink



United States [change] [Canada]

Search

Home | Solutions | Services | Products | Support & downloads | My IBM | Welcome Mr. Alfred Christensen [Not you?] [IBM Sign out]

Resource Link
Site search
Planning
Education
Library
Fixes
Problem solving
Services
Tools
Customer Initiated Upgrade
Feedback

IBM Systems > System z > Resource Link > Library >

Open Systems Adapter (OSA) Library

OSA-Express Direct SNMP Support

Driver support level	MIB module	MIBS.DATA file
OSA-Express3	osaexp3.mib (186KB)	osaexp3.mibs.data (23KB)
IBM System z9, z990, and z890 at May 2004 OSA-Express LIC or above	osa10gig.mib (152KB)	osa10gig.mibs.data (18KB)
IBM z990 and z890 with OSA-Express LIC prior to May 2004	osa333c.mib (122KB)	osa333c.mibs.data (15KB)
IBM z900 and z800 at minimum Driver 3G with OSA-Express LIC level 3.33	osa333c.mib (122KB)	osa333c.mibs.data (15KB)
IBM z900 and z800 Driver 3G with OSA-Express LIC level 3.0A at MCLs: J11204.007 and J11204.008	osa030c.mib (110KB)	osa030c.mibs.data (14KB)

Related OSA information

→ OSA-Express for zSeries and S/390
→ OSA-2 to OSA-Express Migration
→ OSA Support Facility (OSA/SF) for VM

Let IBM help you with your next IT project

Download our new eCatalog and discover what IBM IT Services can do for you.

→ Learn more

How to enable the OSA Direct SNMP support

- Configure and start the z/OS CS SNMP daemon, if not already done so
  - See the z/OS CS documentation for doing this
- Download the relevant OSA mibs.data file and append it to your /etc/mibs.data file
- Configure and start the IOBSNMP address space
  - Sample JCL in hlq.SEZAINST(IOBSNMP)

```
//IOBSNMP  PROC P='-s TCPCS'
//IOBSNMP  EXEC PGM=IOBSNMP,TIME=1440,REGION=4096K,DYNAMNBR=5,
//         PARM='&P.'
//SYSPRINT DD  SYSOUT=*
//SYSUDUMP DD  SYSOUT=*
//*
//* The options available for the PARMS referenced by P are
//* -d level   - turn on specified debugging
//*              0 - no tracing
//*              1 - minimal tracing
//*              2 - maximum tracing
//*             >2 - maximum tracing plus SNMP traces
//* -c community   - use specified community name
//* -p port number - use specified port number
//* -s stack       - send request to specified stack
//*
//* Defaults:  -d 0 -c public  -p 161  -s Default_stack
```

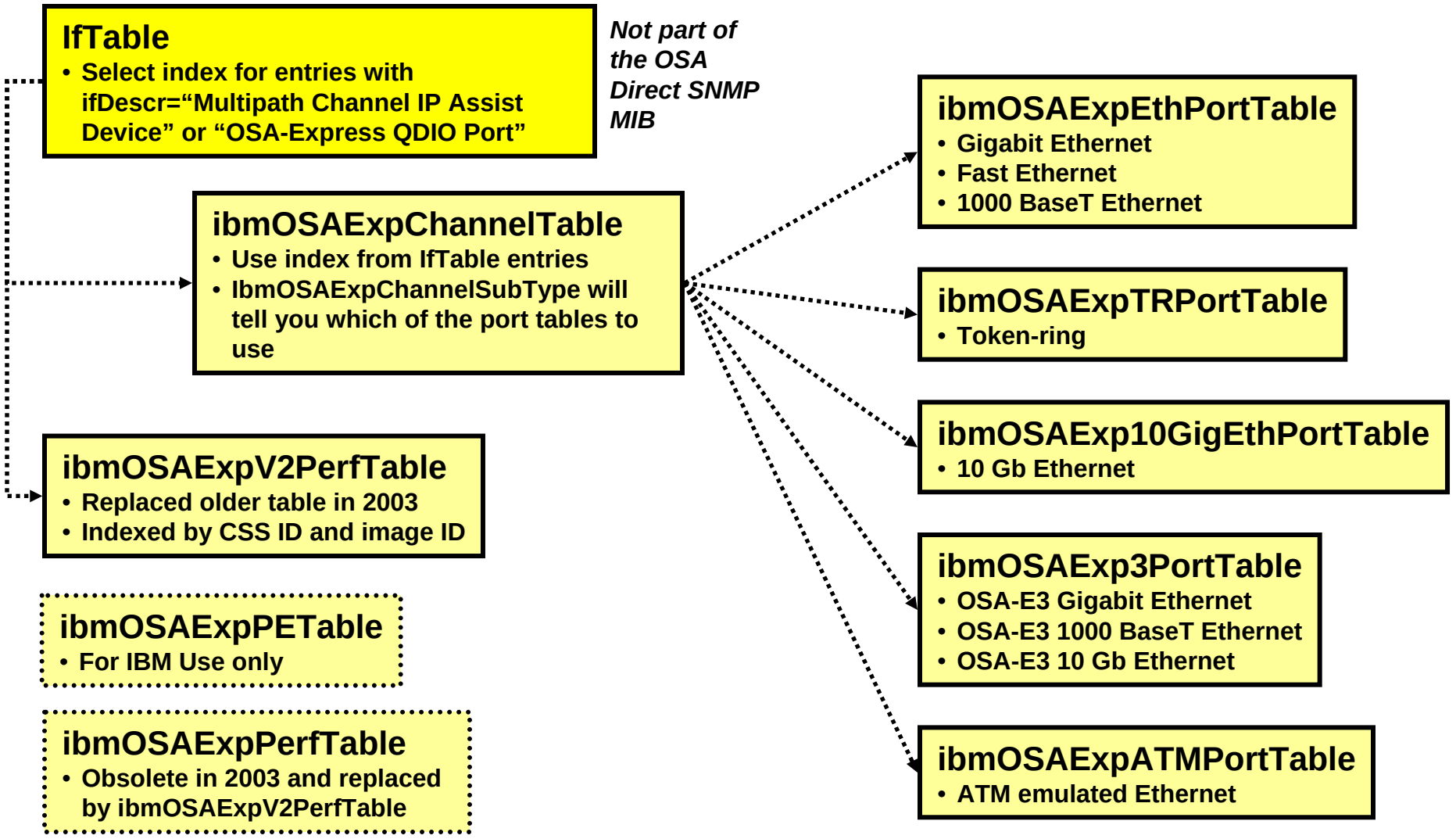
This is the SNMPv1 community name (password). Default is 'public'. Can be set in your SNMP daemon PWSRC file.

This is the SNMP daemon UDP port number to send SNMP requests to. Default is 161.

```
OSNMPD   00000022 UDP
  Local Socket:   ::..161
  Foreign Socket: *..*
```

The name of the TCP/IP stack address space to use for communication with the SNMP daemon and the OSA adapter.

OSA Direct SNMP MIB structure



Navigating to port-specific information

1 **snmp –v walk ifTable**

```
. . . . .
ifIndex.19 = 19
ifIndex.20 = 20
ifDescr.1 = Loopback Device
ifDescr.2 = Loopback
ifDescr.3 = Loopback IPv6
ifDescr.5 = Virtual IP Address IPv6
ifDescr.6 = Virtual IP Address Device
ifDescr.7 = Virtual IP Address Link
ifDescr.8 = Multipath Channel Point-to-Point Device
ifDescr.9 = Multipath Channel Point-to-Point
ifDescr.10 = LCS Device
ifDescr.11 = Token Ring
ifDescr.12 = Multipath Channel IP Assist Device
ifDescr.13 = IP Assist QDIO Ethernet
ifDescr.14 = IP Assist QDIO Ethernet IPv6
ifDescr.17 = Virtual IP Address Device
ifDescr.18 = Virtual IP Address Link
ifDescr.19 = Virtual IP Address Device
ifDescr.20 = Virtual IP Address Link
ifType.1 = 53
ifType.2 = 24
. . . . .
```

2 **snmp –v get ifIndex.12**

```
ifIndex.12 = 12
```

3 **snmp –v walk ibmOSAExpChannelTable**

```
ibmOSAExpChannelNumber.12 = '00d0'h
ibmOSAExpChannelType.12 = 17
ibmOSAExpChannelHdwLevel.12 = 5
ibmOSAExpChannelSubType.12 = 177
ibmOSAExpChannelShared.12 = 1
ibmOSAExpChannelNodeDesc.12 =
'200006d0f0f0f1f7f3f0f0f0f5c9c2d4f0f2f0f0f0f
0f0f0f0c2c5f1f5c5d000'h
ibmOSAExpChannelProcCodeLevel.12 = '0751'h
ibmOSAExpChannelPCIBusUtil1Min.12 = 0
ibmOSAExpChannelProcUtil1Min.12 = 0
ibmOSAExpChannelPCIBusUtil5Min.12 = 0
ibmOSAExpChannelProcUtil5Min.12 = 0
ibmOSAExpChannelPCIBusUtilHour.12 = 0
ibmOSAExpChannelProcUtilHour.12 = 0
```

The PCI Bus and Processor utilization data are for the entire CHPID; not just for this LPAR.

The ibmOSAExpChannelSubType value of 177 tells us to use the OSA-Express3 port table (ibmOSAExp3PortTable)

Interface index 12 from the TCP/IP stack is reused to select entries in the ibmOSAExpChannelTable and later in the ibmOSAExp3PortTable.

OSA-E3 port-specific information

snmp –v walk ibmOSAExp3PortTable

```
ibmOsaExp3PortNumber.12 = 0
ibmOsaExp3PortType.12 = 177
ibmOsaExp3LanTrafficState.12 = 4
ibmOsaExp3ServiceMode.12 = 0
ibmOsaExp3DisabledStatus.12 = '0000'h
ibmOsaExp3ConfigName.12 = IBM Default Configuration File
ibmOsaExp3ConfigSpeedMode.12 = 0
ibmOsaExp3ActiveSpeedMode.12 = 6
ibmOsaExp3MacAddrActive.12 = '00145e74e0f0'h
ibmOsaExp3MacAddrBurntIn.12 = '00145e74e0f0'h
ibmOsaExp3PortName.12 = OSAQDIO4
ibmOsaExp3TotalPacketsXmit.12 = 1148636
ibmOsaExp3TotalPacketsRecv.12 = 1073478
ibmOsaExp3GoodPacketsXmit.12 = 1148636
ibmOsaExp3GoodPacketsRecv.12 = 1073478
ibmOsaExp3Packet64Xmit.12 = 165369
ibmOsaExp3Packet65to127Xmit.12 = 144365
ibmOsaExp3Packet128to255Xmit.12 = 51352
ibmOsaExp3Packet256to511Xmit.12 = 8292
. . . . . .
```

These traffic numbers represent the aggregate workload of all the LPARs and TCP/IP stacks that share this OSA port.

A single TCP/IP stack's traffic numbers can be found in a netstat devlinks report.

What can you do with all this information?

You can obviously buy a MIB browser to look at it, but you need something that is intelligent enough to translate the many enumerations into meaningful text.

There are intelligent software solutions that can look at this data for you.

You can also, relatively easily write some simple REXX logic to go and pull it up and print it out the way you want it.

This is a small example of such an approach.

```
OSA SNMP data for: OSAQDIO4 on channel: 00d0 w. MAC address: 00145e74e0f0

Channel   - Number ............. 00d0
=======   - Type ................ OSD
          - Subtype ............. OSA-E3 1000BaseT Ethernet
          - Hardware level ...... OSA-Express3 4.00
          - Share mode .......... Shared channel
Processor - Code level .......... 7.51
========= - Util last minute ....   0.00%
          - Util last 5 minutes .   0.00%
          - Util last hour ......   0.00%
PCI bus   - Util last minute ....   3.00%
=======   - Util last 5 minutes .   3.00%
          - Util last hour ......   3.00%
Port      - Port number ......... 0
====      - Port type ........... OSA-E3 1000BaseT Ethernet
          - Traffic state ....... Enabled
          - Service mode ........ Not in service mode
          - Configuration name .. IBM DEFAULT CONFIGURATION FILE
          - Configured speed .... Auto negotiate
          - Active speed ........ 1000 Mb Full Duplex
          - Burned-in MAC ....... 00145e74e0f0
          - Active MAC .......... 00145e74e0f0
          - Port name ........... OSAQDIO4
```

Detailed traffic data for an OSA port

```
Transmit  - Total packets.............    11368821 100.00%
          - Good packets..............    11368821 100.00%
          - Broadcast packets.........        1331   0.01%
          - Multicast packets.........      365247   3.21%
          - Size: 64 bytes or less...      2860015  25.16%
          - Size: 65 to 127 bytes....      2290829  20.15%
          - Size: 128 to 255 bytes...       554968   4.88%
          - Size: 256 to 511 bytes...        75555   0.66%
          - Size: 512 to 1023 bytes..       164339   1.45%
          - Size: 1024 bytes or more.      5423115  47.70%
Receive   - Total packets.............    15265498 100.00%
          - Good packets..............    15265498 100.00%
          - Broadcast packets.........      140888   0.92%
          - Multicast packets.........     1716694  11.25%
          - Size: 64 bytes or less...      1907217  12.49%
          - Size: 65 to 127 bytes....      3549608  23.25%
          - Size: 128 to 255 bytes...       174752   1.14%
          - Size: 256 to 511 bytes...       284678   1.86%
          - Size: 512 to 1023 bytes..        79139   0.52%
          - Size: 1024 bytes or more.      9270104  60.73%
```

Detailed LAN error statistics for an OSA port

```
Errors    - Alignment errors.........          0
          - CRC errors...............          0
          - Missed packets...........          0
          - Single collisions........          0
          - Multiple collisions......          0
          - Excessive collisions.....          0
          - Late collisions..........          0
          - Deferred.................          0
          - Sequence errors..........          0
          - No receive buffer........          0
          - Receive length error.....          0
          - XON transmitted..........          0
          - XON received.............          0
          - XON transmitted..........          0
          - XOFF received............          0
          - Receive jabber...........          0
          - Receive undersize........          0
          - Receive oversize.........          0
          - Receive fragment.........          0
```

Getting the most out of your OSA adapter with z/OS Communications Server

Appendix A: OSA-Express3 dual port IOCP, TRLE, and z/OS CS Interface definitions

OSA-Express3 IOCP definitions

```
CHPID PATH=(CSS(0,1),FD),SHARED,                                    *
      PCHID=581,TYPE=OSD

CNTLUNIT CUNUMBR=2FD0,PATH=((CSS(0),FD),(CSS(1),FD)),UNIT=OSA

IODEVICE  ADDRESS=(2FD0,14),CUNUMBR=2FD0,UNIT=OSA,                  *
      UNITADD=00
IODEVICE  ADDRESS=(2FDE,1),CUNUMBR=2FD0,UNIT=OSAD,                  *
      UNITADD=FE
```

```
10.37.39  d m=chp(fd)
      10.37.39  IEE174I 10.37.39 DISPLAY M 825                                  C
      CHPID FD:  TYPE=11, DESC=OSA DIRECT EXPRESS, ONLINE
      DEVICE STATUS FOR CHANNEL PATH FD
           0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F
      02FD +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  .
      SWITCH DEVICE NUMBER = NONE
      PHYSICAL CHANNEL ID = 0581
      ************************ SYMBOL EXPLANATIONS ************************
      + ONLINE    @ PATH NOT VALIDATED   - OFFLINE    . DOES NOT EXIST
      * PHYSICALLY ONLINE   $ PATH NOT OPERATIONAL
```

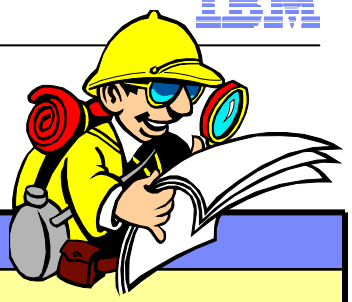
OSA Express3 TRLE definitions

```
O3ETHB0T TRLE  LNCTL=MPC,                                  X
               READ=(2FD0),                                X
               WRITE=(2FD1),                               X
               MPCLEVEL=QDIO,                              X
               DATAPATH=(2FD2,2FD3,2FD4,2FD5),             X
               PORTNAME=(O3ETHB0P),                        X
               PORTNUM=(0)

O3ETHB1T TRLE  LNCTL=MPC,                                  X
               READ=(2FD6),                                X
               WRITE=(2FD7),                               X
               MPCLEVEL=QDIO,                              X
               DATAPATH=(2FD8,2FD9,2FDA,2FDB),             X
               PORTNAME=(O3ETHB1P),                        X
               PORTNUM=(1)
```

OSA Express3 TCP/IP Interface definitions

```
INTERFACE O3ETHB0 DEFINE IPAQENET
 PORTNAME O3ETHB0P
 IPADDR 16.11.16.105/20
 SOURCEVIPAINT LGEVIPA1
 MTU 1500
 VLANID 3
 READSTORAGE GLOBAL
 INBPERF BALANCED
 IPBCAST
 MONSYSPLEX
 DYNVLANREG
 OLM
 VMAC 0204100B1069 ROUTEALL

INTERFACE O3ETHB1 DEFINE IPAQENET
 PORTNAME O3ETHB1P
 IPADDR 16.11.17.105/20
 SOURCEVIPAINT LGEVIPA1
 MTU 1500
 VLANID 3
 READSTORAGE GLOBAL
 INBPERF BALANCED
 IPBCAST
 MONSYSPLEX
 DYNVLANREG
 OLM
 VMAC 0204100B1169 ROUTEALL
```

For more information



URL	Content
http://www.twitter.com/IBM_Commserver	IBM Communications Server Twitter Feed
http://www.facebook.com/IBMCommserver	IBM Communications Server Facebook Fan Page
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.ibm.com/software/network/commserver/z_lin/	IBM Communications Server for Linux on System z
http://www.ibm.com/software/network/ccl/	IBM Communication Controller for Linux on System z
http://www.ibm.com/software/network/commserver/library/	IBM Communications Server library
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/atsmastr.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

For pleasant reading ….