#SHAREorg

zEnterprise System z/OS IEDN network design and implementation (Part 2)

Gus Kassimis - kassimis@us.ibm.com
IBM Enterprise Networking Solutions
Raleigh, NC, USA

Session: 11335
Tuesday, August 7, 2012: 3:00 PM-4:00 PM

SHARE in Anaheim 2012

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.

zEnterprise System - z/OS IEDN network design and implementation (Part 2)

Session number:	11335
Date and time:	Tuesday, August 7, 2012: 3:00 PM-4:00 PM
Location:	Platinum Ballroom Salon 9 (Anaheim Marriott Hotel)
Program:	Communications Infrastructure
Project:	Communications Server
Track:	Network Support and Management and z/OS Systems Programming
Classification:	Technical
Speaker:	Gus Kassimis, IBM
Abstract:	There are many technologies included in the z/OS Communications Server that enable you to design and implement highly available and scalable z/OS Sysplex networking scenarios, and most of you have been using those technologies for many years. How do you extend those capabilities to the zEnterprise environment, or more specifically, to the Inter-Ensemble Data Network (IEDN)? When a z/OS system is part of a z/OS Sysplex and needs to be connected to both a remote network and to the IEDN, some special considerations must be used when designing for high availability and workload management. This session will analyze those considerations and discuss a few selected network topologies to illustrate the design and implementation considerations. The session will cover topologies that include applications deployed on Power and X86 blades and also optimizers in the zBX, such as the IBM WebSphere DataPower Integration Appliance XI50 for zEnterprise (DataPower XI50z).

Related zEnterprise Networking Sessions

(11334): zEnterprise System - Network Architecture and Virtualization Overview (Part 1)	Tuesday, August 7, 2012: 1:30 PM-2:30 PM Platinum Ballroom Salon 9 (Anaheim Marriott Hotel) Speaker: Gus Kassimis(IBM Corporation)
(11335): zEnterprise System - z/OS IEDN network design and implementation (Part 2)	***Tuesday, August 7, 2012: 3:00 PM-4:00 PM*** ***Platinum Ballroom Salon 9 (Anaheim Marriott Hotel)*** ***Speakers: Gus Kassimis (IBM Corporation)***
(11901): zEnterprise System - Secure Networking with the zEnterprise Ensemble – (Part 3)	Tuesday, August 7, 2012: 4:30 PM-5:30 PM Platinum Ballroom Salon 9 (Anaheim Marriott Hotel) Speaker: Gwen Dente (IBM Corporation)
(11154): Connectivity to the zBX and Routing in the zEnterprise Ensemble	Monday, August 6, 2012: 4:30 PM-5:45 PM Grand Ballroom Salon A (Anaheim Marriott Hotel) Speaker: Gwen Dente (IBM Corporation)
(11350): zEnterprise Networking Lessons Learned	Friday, August 10, 2012: 9:30 AM-10:30 AM Platinum Ballroom Salon 9 (Anaheim Marriott Hotel) Speaker: Tom Cosenza (IBM Corporation)

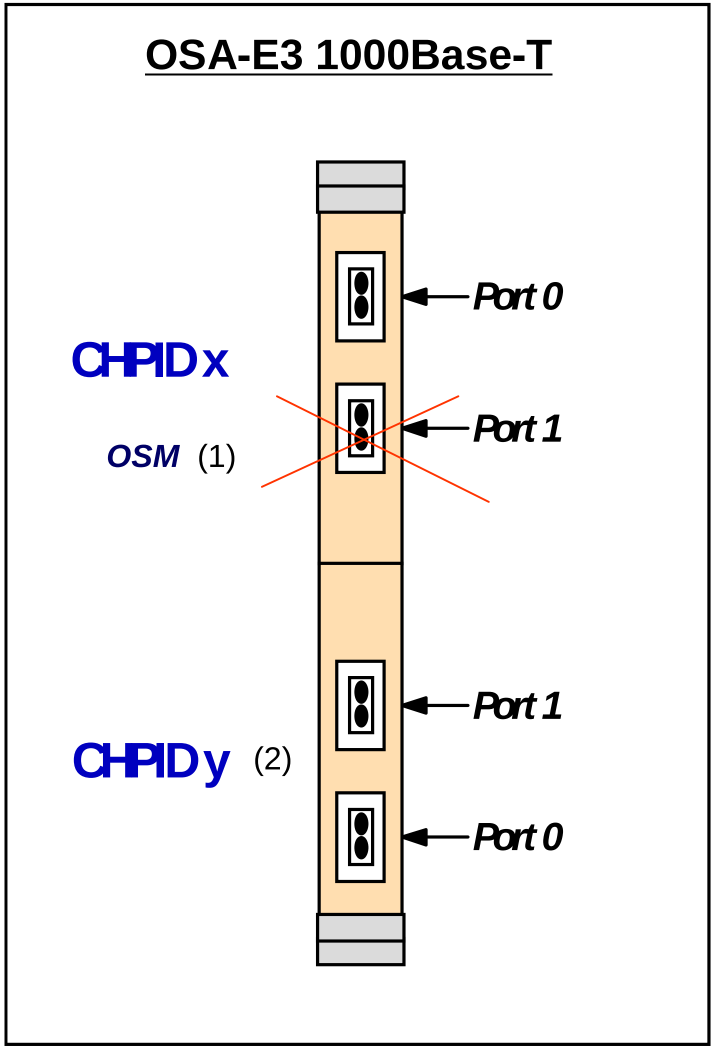
Agenda

- z/OS configuration for zEnterprise
- Basics of interface recovery and use of dynamic VIPAs on the IEDN
- IEDN implementation scenarios:
  1. New multi-tier workload
  2. Extending existing sysplex workload with IEDN workload
  3. DataPower and Sysplex Distributor
- IEDN enabled HiperSockets and configuration considerations

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.

z/OS configuration for zEnterprise

The zEnterprise Management Network 1000Base-T OSA configured as CHPID type OSM



IOCDS

```
CHPID PCHID=191,PATH=(CSS(0,1,2,3),23),              *
TYPE=OSM,CHPARM=01,SHARED,   …

CNTLUNIT CUNUMBR=0910,PATH=((CSS(0),23)),           *
UNIT=OSM

IODEVICE ADDRESS=(0910,15),CUNUMBR=(0910),  *
UNIT=OSA,UNITADD=00,                        *
MODEL=M,DYNAMIC=YES,LOCANY=YES
```

z/OS Operating System Definitions

Not Applicable

These Connections are dynamically created when an Ensemble is defined.

z/OS Communications Server and OSM connectivity

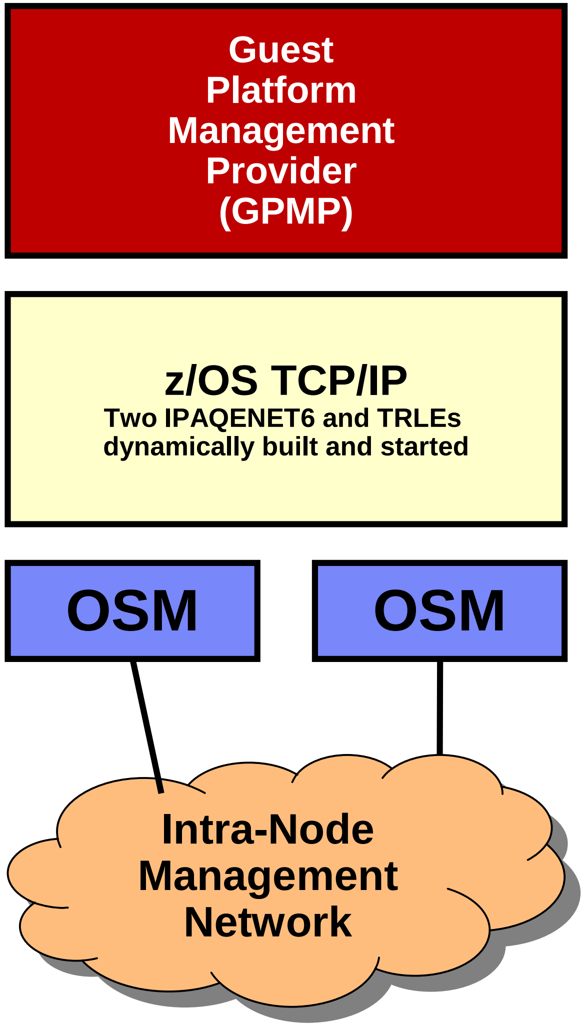
- z/OS LPAR must participate in the ensemble
  - New VTAM start option: ENSEMBLE=YES
    - Required for both OSM and OSX connectivity

```
ENSEMBLE=YES,              LPAR is part of zEnterprise ENSEMBLE
```

- LPAR must be IPv6-enabled for OSM connectivity
  - BPXPRMxx PARMLIB updates
- Two IPAQENET6 interface definitions are dynamically generated and started
  - If OSM CHPIDs are defined to the z/OS image, the two CHPIDs with the lowest device numbers are assigned to these interfaces
- TRLEs dynamically generated if connectivity allowed and CHPIDs found
  - Only port 0 supported
- IPv6 link-local address only
  - No globally unique IPv6 addresses needed

z/OS Communications Server and OSM connectivity (cont)

- Uses VLAN in access mode
  - Switch handles VLAN tagging, stack unaware
- Not reported to OMPROUTE
- Cannot add static or dynamic routes
- Supports stop, start, packet trace, and OSA NTA
- Only applications permitted to EZB.OSM.sysname.tcpname can communicate over OSM interfaces
  - The Guest Platform Management Provider (GPMP) is the only application that needs to



Steps to enable the intranode management network

1. Authorize the management application to the EZB.OSM.sysname.tcpname resource.
   – To send or receive data over an OSM interface, an application must have READ authorization to the EZB.OSM.sysname.tcpname resource. If used on this image, authorize the application to this resource.

2. Reserve the UDP port that the platform management application is to use to listen for multicast traffic over the intranode management network.

3. Authorize any user IDs to this resource that might issue diagnostic commands, such as Ping and Traceroute, over OSM interfaces to verify connectivity.

4. If you enable IP security for IPv6, you can configure a security class for IP filtering that applies to all OSM interfaces.
   – Use the OSMSECCLASS parameter on the IPCONFIG6 statement. This enables you to configure filter rules for traffic over the EZ6OSM01 and EZ6OSM02 interfaces.

5. If the multicast address that is used by the platform management application is configured into a network access zone, then give the user ID for this application read permission to the resource profile for that zone.

How bad is it enabling IPv6?

- Add a NETWORK stmt. for AF_INET6 to your BPXPRMxx PARMLIB member
- No changes needed to your TCP/IP Profile
  - Unless you want to exploit and test specific IPv6 features

```
FILESYSTYPE TYPE(INET) ENTRYPOINT(EZBPFINI)
 NETWORK DOMAINNAME(AF_INET)
         DOMAINNUMBER(2)
         MAXSOCKETS(2000)
         TYPE(INET)
 NETWORK DOMAINNAME(AF_INET6)
         DOMAINNUMBER(19)
         MAXSOCKETS(3000)
         TYPE(INET)
```

AF_INET

AF_INET6

- Testing needed in the network management area
  - All Netstat reports will use the LONG format
    - Also when reporting on IPv4-only activity
  - Home-written Netstat “scraping” logic will need to be changed
  - Network management products may fail if they are not prepared for IPv6 addresses

```
MVS TCP/IP NETSTAT CS V1R12       TCPIP Name: TCPCS           13:02:52
User Id  Conn      State
-------  ----      -----
MYINETD1 00000022 Listen
  Local Socket:   9.42.130.98..23
  Foreign Socket: 0.0.0.0..0
TN3270A  0000004D Establsh
  Local Socket:   ::ffff:9.42.105.45..23
  Foreign Socket: ::ffff:9.76.144.213..4211
  Application Data:  EZBTNSRV TCPABC80 TSO10001 ET B
TN3270A  0000003F Listen
  Local Socket:   ::..23
  Foreign Socket: ::..0
  Application Data:  EZBTNSRV LISTENER
```

AF_INET socket

AF_INET6 sockets

Netstat devlinks example of an OSM interface

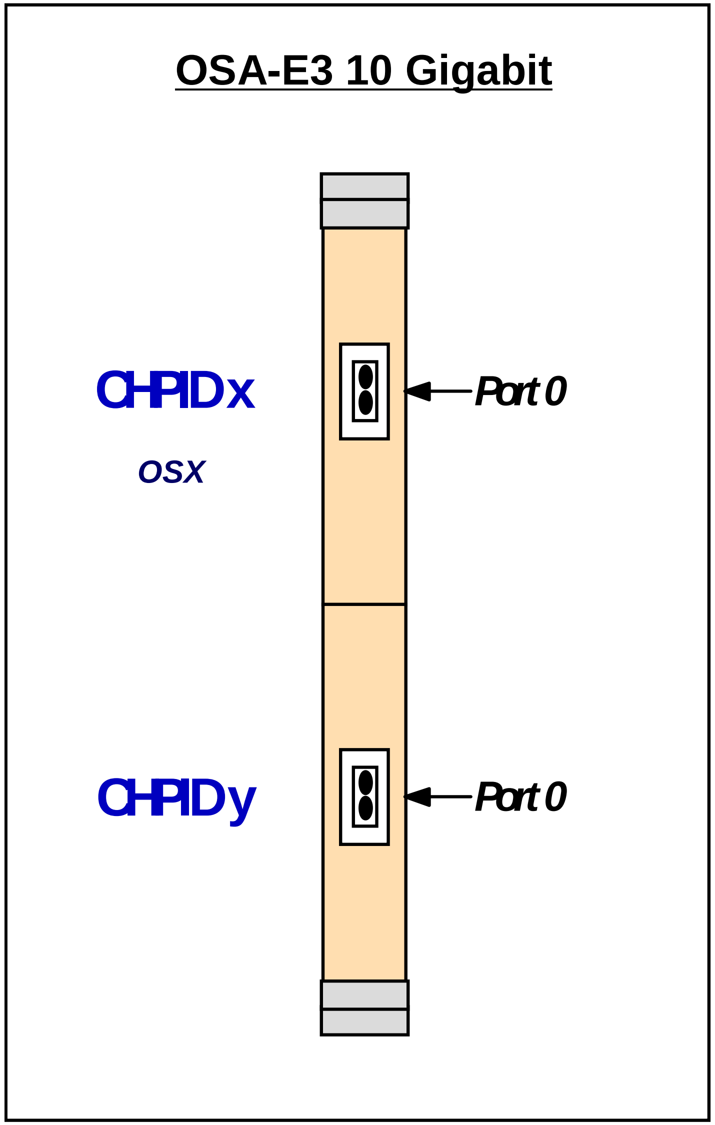
```
IntfName: EZ6OSM01          IntfType: IPAQENET6  IntfStatus: Ready
    PortName: IUTMP0F2  Datapath: BE22     DatapathStatus: Ready
    CHIPDType: OSM
    QueSize: 0    Speed: 0000001000
    VMACAddr: 02006FEB8363   VMACOrigin: OSA    VMACRouter: All
    DupAddrDet: 1
    CfgMtu: None                     ActMtu: 1500
    VLANid: None                     VLANpriority: Disabled
    ReadStorage: GLOBAL (4096K)
    InbPerf: Dynamic
      WorkloadQueueing: No
    ChecksumOffload: No              SegmentationOffload: No
    SecClass: 255                    MonSysplex: No
    Isolate: Yes                     OptLatencyMode: No
    TempPrefix: None
 Multicast Specific:
    Multicast Capability: Yes
    Group:     ff02::1:ffeb:8363
      RefCnt:  0000000001  SrcFltMd: Exclude
      SrcAddr: None
(more multicast group memeberships)
 Interface Statistics:
    BytesIn                           = 4256
    Inbound Packets                   = 29
    Inbound Packets In Error          = 9
    Inbound Packets Discarded         = 0
    Inbound Packets With No Protocol  = 0
    BytesOut                          = 958
    Outbound Packets                  = 9
    Outbound Packets In Error         = 0
    Outbound Packets Discarded        = 0
```

CHPIDType OSM

VMAC

INBPERF is Dynamic

The zEnterprise Internal Data Network 10 Gigabit OSA configured as CHPID type OSX (z/OS)



IOCDS

CHPID PCHID=5E1,PATH=(CSS(0,1,2,3),2F), *
TYPE=OSX,SHARED, …

CNTLUNIT CUNUMBR=09F0,PATH=((CSS(0),2F)), *
UNIT=OSX

IODEVICE ADDRESS=(09F0,15),CUNUMBR=(091F),*
UNIT=OSA,UNITADD=00, *
MODEL=X,DYNAMIC=YES,LOCANY=YES

Operating System Definitions (z/OS)

VTAM Definitions

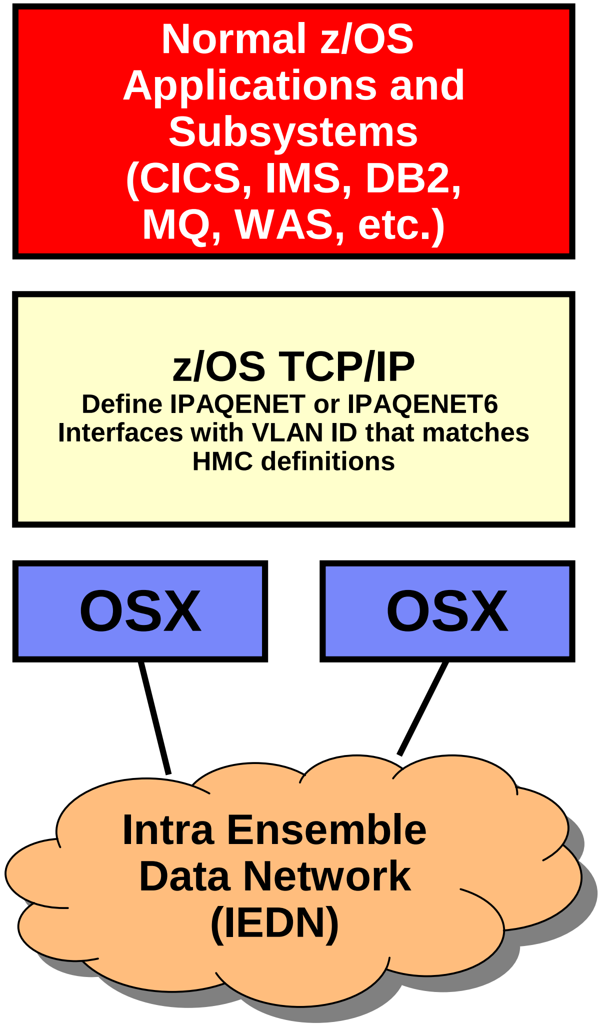
- Dynamic TRLEs
- Manually Defined TRLEs

TCP/IP Definitions

- INTERFACE IPAQENET
- INTERFACE IPAQENET6

z/OS Communications Server and OSX connectivity

- Configure with INTERFACE statement
  - IPAQENET and IPAQENET6
- Either specify CHPID
  - Dynamically created TRLE similar to HiperSockets
- Or configure TRLE and point to it
  - Useful in VM guest LAN environment where CHPID is unpredictable
- Always uses VLAN in trunk mode
  - VLANID required and must be authorized at HMC
    - If not authorized, OSA activation fails
    - Note: in z/OS V1R13 up to 32 VLANs can be defined per OSX interface (prior to V1R13 the limit was 8)
- Prevents IP forwarding from OSX ⇨ OSX
  - Sysplex distributor forwarding is allowed when using VIPAROUTE



z/OS Communications Server and OSX connectivity

- Supports stop, start, packet trace, and OSA NTA
- To prevent external traffic from being routed to/from the OSX VLAN
  - Define OSX as INTERFACE or IPV6_INTERFACE
  - Do not enable IMPORT_DIRECT_ROUTES function
  - Alternatively, do not define the OSX interfaces to OMPROUTE and tell OMPROUTE to ignore undefined interfaces
- To allow external traffic to be routed to/from the OSX VLAN
  - Define OSX as OSPF_INTERFACE or IPV6_OSPF_INTERFACE
  - Configure a non-0 value for ROUTER_PRIORITY
- If you are already using OMPROUTE and OSPF, defining the IEDN interfaces to OSPF may be what you prefer
  - Routing to/from OSX interfaces can still be controlled via IPSec filter rules

IEDN OSX IOCP and TCP/IP sample definitions

IOCP

```
CHPID PATH=(CSS(0,1),D3),SHARED,                                  *
      PCHID=131,TYPE=OSX
CNTLUNIT CUNUMBR=BF30,PATH=((CSS(0),D3),(CSS(1),D3)),UNIT=OSA
IODEVICE  ADDRESS=(BF30,14),CUNUMBR=BF30,UNIT=OSA,                *
      NOTPART=((CSS(1),RALHCD)),                                  *
      UNITADD=00
```

CHPIDTYPE OSX

CHPID D3

VLANID 85

TCP/IP

```
INTERFACE O3OSXA0 DEFINE IPAQENET
 CHPIDTYPE OSX
 CHPID D3
 IPADDR 16.11.160.108/21
 SOURCEVIPAINT LFRVIPA1
 MTU 8992
 VLANID 85
 READSTORAGE GLOBAL
 INBPERF DYNAMIC WORKLOADQ
 NOISOLATE
 IPBCAST
 MONSYSPLEX
 DYNVLANREG
 NOOLM
 VMAC
```

Virtual MAC address

OMPROUTE

```
OSPF_Interface
  Name                       = O3OSXA0
  IP_Address                 = 16.11.160.108
  Subnet_Mask                = 255.255.248.0
  MTU                        = 9000
  Hello_Interval             = 10
  Dead_Router_Interval       = 40
  Retransmission_Interval    = 20
  Cost0                      = 100
  Attaches_To_Area           = 11.11.11.11
```

Netstat devlinks example of an OSX interface

```
IntfName: O3OSXA0          IntfType: IPAQENET    IntfStatus: Ready
    PortName: IUTXP0D3  Datapath: BF32     DatapathStatus: Ready
    CHPIDType: OSX      CHPID: D3
    Speed: 0000010000
    IpBroadcastCapability: Yes
    VMACAddr: 029BA400024F   VMACOrigin: OSA    VMACRouter: All
    SrcVipaIntf: LFRVIPA1
    ArpOffload: Yes                  ArpOffloadInfo: Yes
    CfgMtu: 8992                     ActMtu: 8992
    IpAddr: 16.11.160.108/21
    VLANid: 85                       VLANpriority: Enabled
    DynVLANRegCfg: Yes               DynVLANRegCap: Yes
    ReadStorage: GLOBAL (4096K)
    InbPerf: Dynamic
      WorkloadQueueing: Yes
    ChecksumOffload: Yes             SegmentationOffload: Yes
    SecClass: 255                    MonSysplex: Yes
    Isolate: No                      OptLatencyMode: No
 Multicast Specific:
    Multicast Capability: Yes
    Group              RefCnt        SrcFltMd
    -----              ------        --------
    224.0.0.6          0000000001    Exclude
      SrcAddr: None
    224.0.0.5          0000000001    Exclude
      SrcAddr: None
    224.0.0.1          0000000001    Exclude
      SrcAddr: None
  Interface Statistics:
    BytesIn                           = 32182373
    Inbound Packets                   = 194050
    Inbound Packets In Error          = 0
    Inbound Packets Discarded         = 0
    Inbound Packets With No Protocol  = 0
    BytesOut                          = 21913806
    Outbound Packets                  = 183213
    Outbound Packets In Error         = 0
    Outbound Packets Discarded        = 0
```

CHPIDType OSX

VMAC

VLANID 85

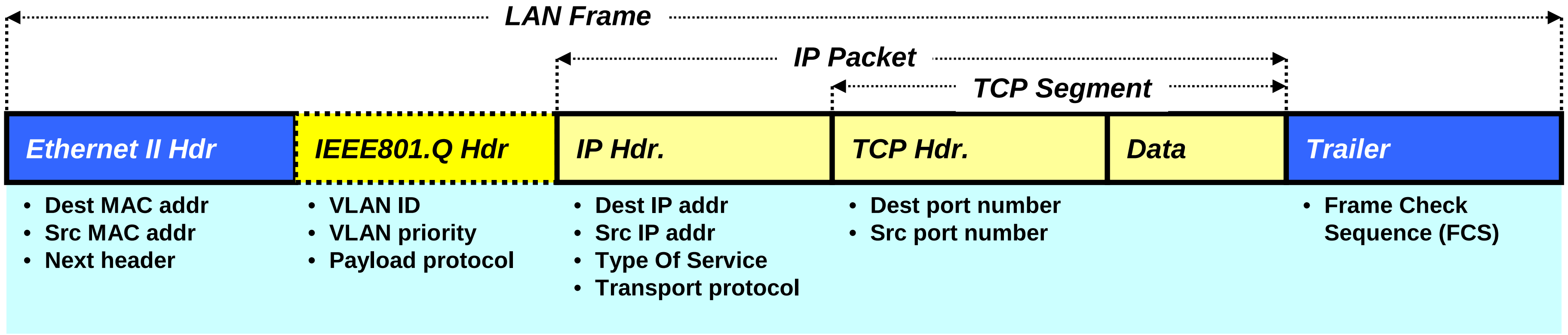
INBPERF Dynamic

zEnterprise networking – z/OS IEDN network design and implementation (Part 2)

Basics of interface recovery and use of dynamic VIPAs on the IEDN

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 physical hardware address
  - A Media Access Control (MAC) address
    - Burned in (world-wide unique by vendors) or alternatively locally administered
      - o Unified Resource Manager administers all MAC addresses in the zBX
- 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 IEEE801.Q header if VLAN technologies are in use
- A frame carries a payload of a specified protocol type, such as ARP, IPv4, IPv6, SNA LLC2, etc.

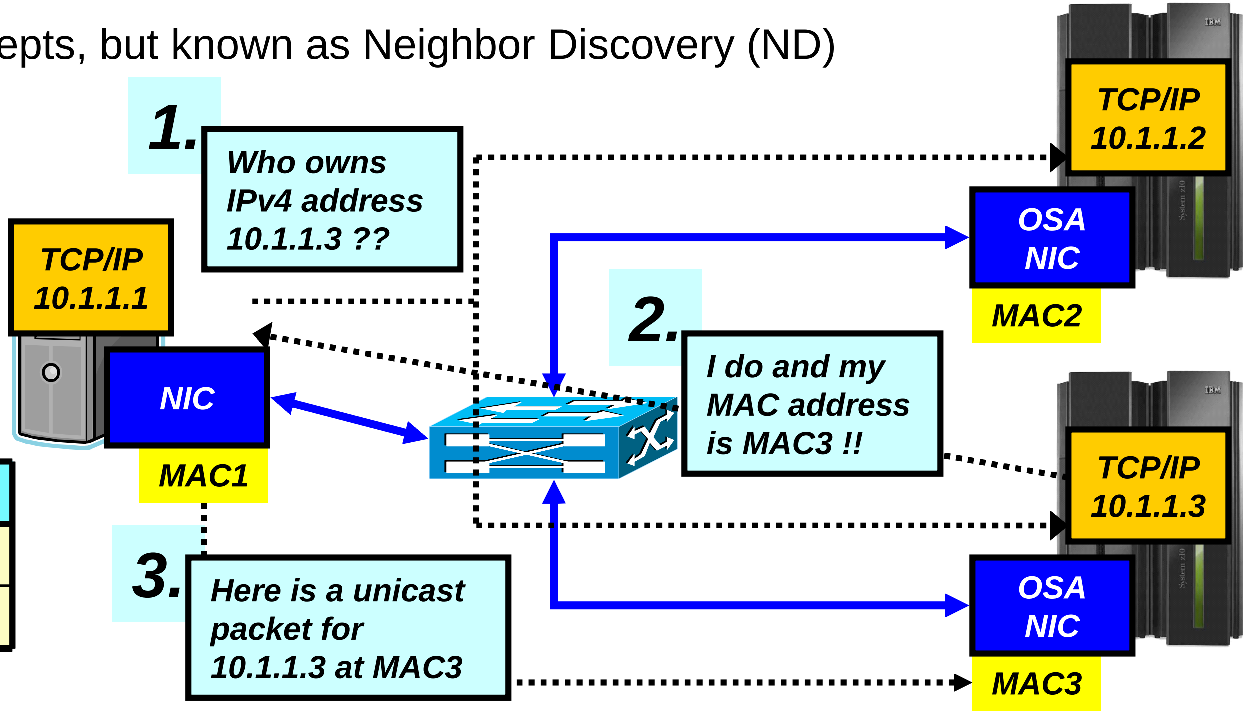


Correlation of IPv4 addresses and MAC addresses on a LAN – Address Resolution Protocol (ARP)

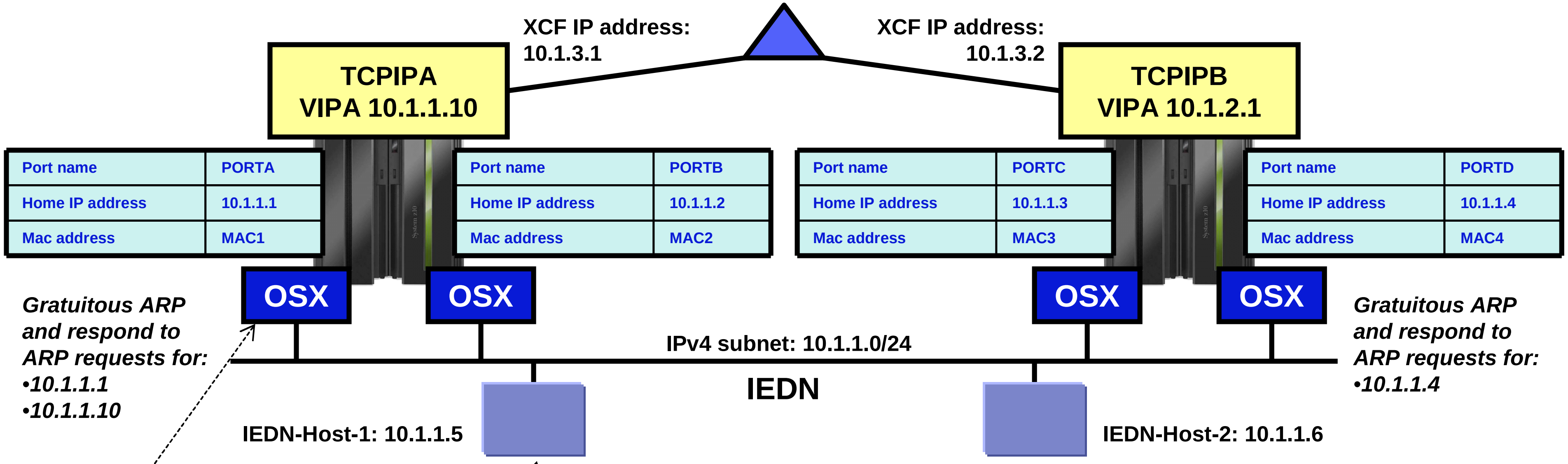
- An IPv4 node uses the ARP protocol to discover the MAC address of another IPv4 address that belongs to the same IPv4 subnet as it does itself.
- ARP requests are broadcasted to all NICs on the LAN
- The one NIC that has a TCP/IP stack with the requested IPv4 address responds directly back to the IPv4 node that sent out the broadcast
- Each IPv4 node maintains a cache of IPv4 addresses and associated MAC addresses on their directly connected LANs
- IPv6 uses similar concepts, but known as Neighbor Discovery (ND)

ARP Cache

IP address	MAC
10.1.1.2	MAC2
10.1.1.3	MAC3



z/OS VIPA address visibility on the IEDN



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

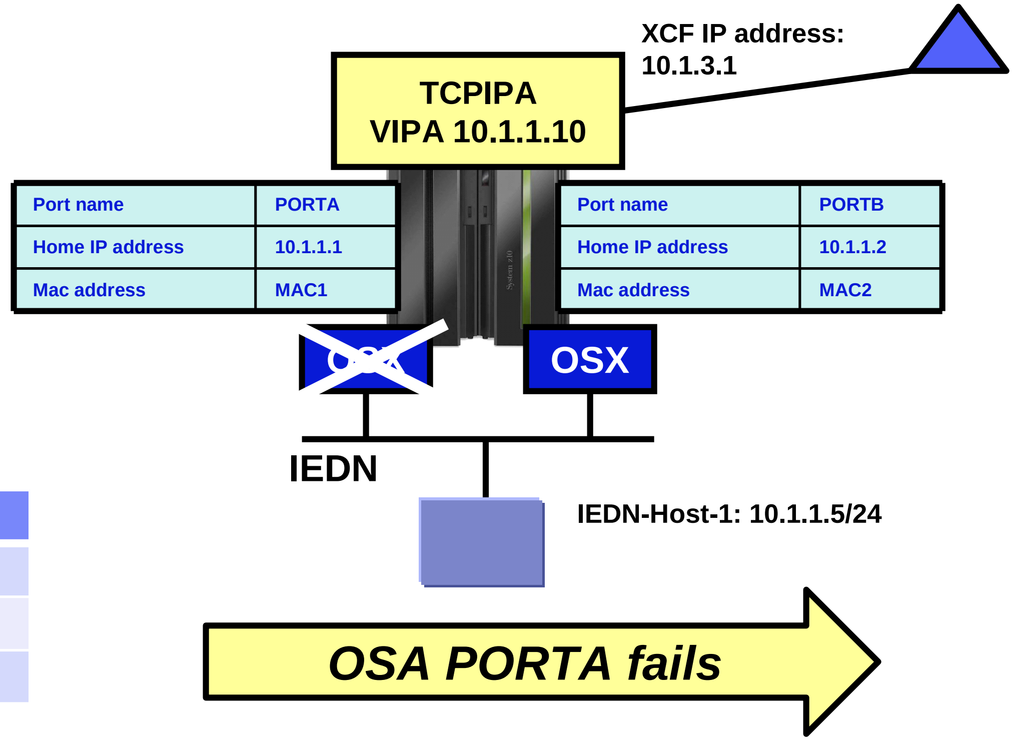
IEDN-Host-1's ARP cache

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

- OSX interfaces must be defined with the INTERFACE statement
- 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.
- OSX interfaces will do gratuitous ARP for the OSA interface IP address and for VIPA addresses that belong to the ***same*** subnet as the OSA interface.

Network connectivity resilience on the IEDN

z/OS TCP/IP supports interface recovery if multiple network interfaces to the same subnet exist.
In this example, both OSA PORTA and PORTB are connected to the IEDN (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 hosts to update their ARP cache.

IEDN-Host-1's ARP cache (before)

IP Address	MAC Address
10.1.1.1	MAC1
10.1.1.2	MAC2
10.1.1.10	MAC1

IEDN-Host-1's ARP cache (after)

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 (crossed out)

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

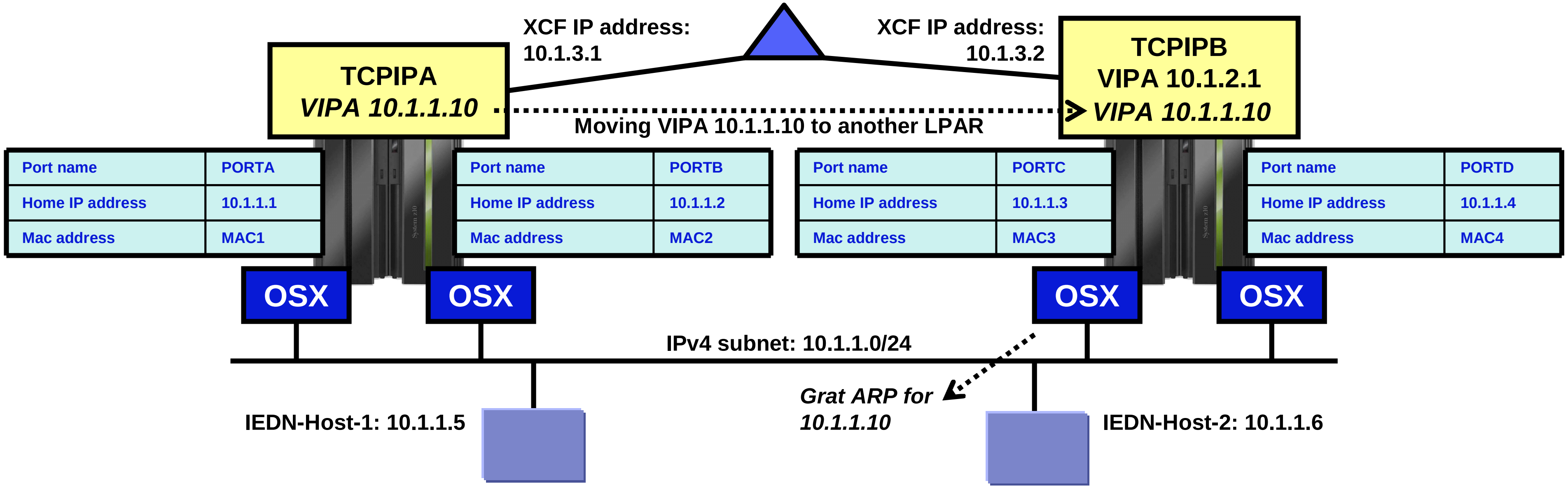
IP Address	ARP Owner
10.1.1.1	Yes
10.1.1.10	Yes
10.1.1.2	Yes
10.1.3.1	No

How do you know which OSA interfaces are on the same subnet and which OSA 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: 00010
  Name                  Status      ArpOwner             VipaOwner
  ----                  ------      --------             ---------
  O3OSXA0               Active      O3OSXA0              Yes
IPv6 LAN Group Summary
LanGroup: 00001
  Name                  Status      NDOwner              VipaOwner
  ----                  ------      -------              ---------
  EZ6OSM02              Active      EZ6OSM02             Yes
  EZ6OSM01              Active      EZ6OSM01             No
```

Dynamic VIPA movement on the IEDN (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

IEDN-Host-1'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

IEDN-Host-1'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

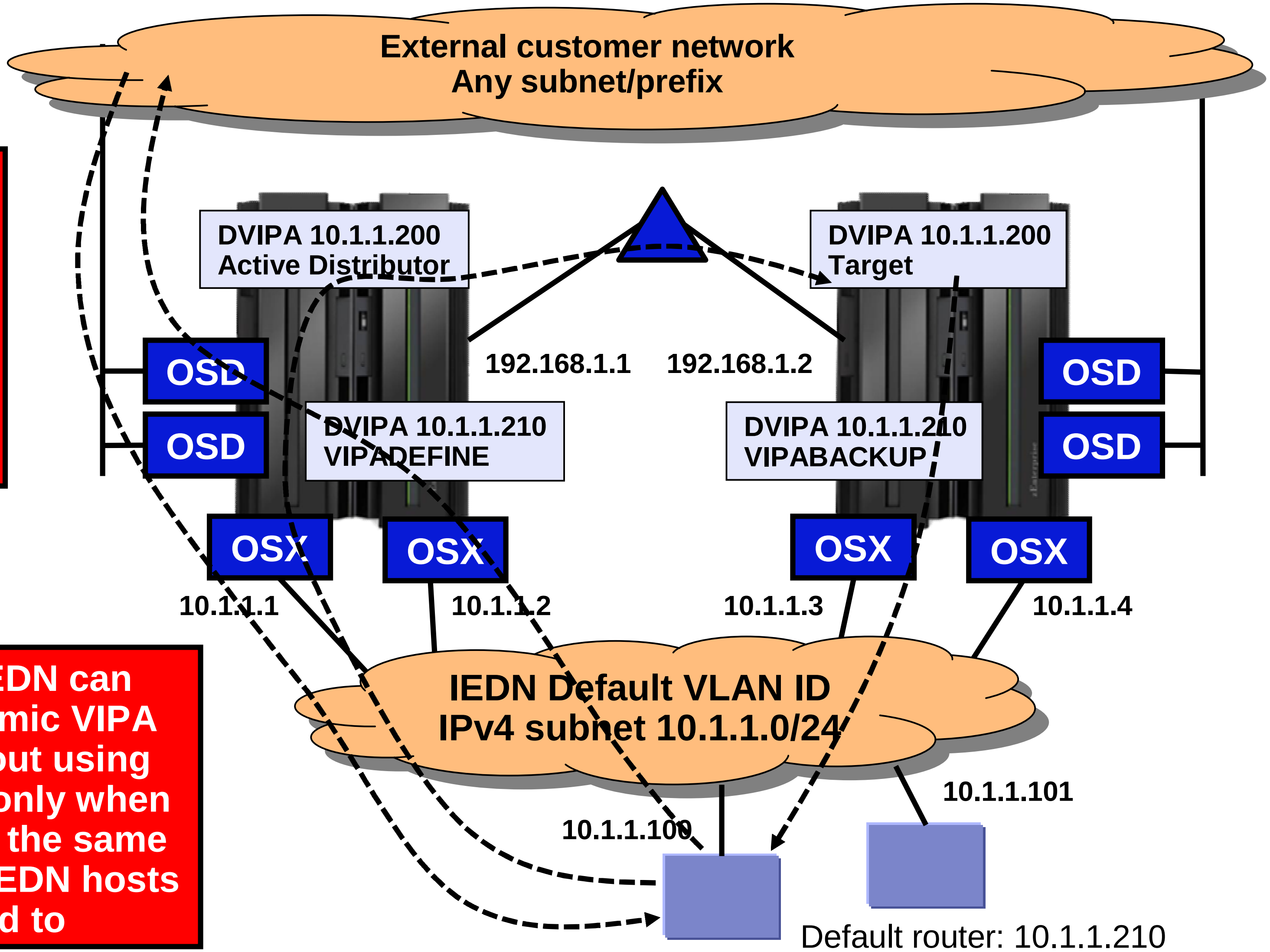
Some of the rules for availability and DVIPA movement in flat networks without dynamic routing – such as on the IEDN

- 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.
  - If multiple VLANs are used on the IEDN, VIPA addresses belonging to one VLAN (one subnet) can in general only be accessed and recovered via that VLAN
  - A default router that does know about all DVIPAs, can be used to reach VIPAs on other VLANs (subnets)
- 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 – part of the zEnterprise infrastructure
  - Redundant OSA adapters in each LPAR – always use at least two OSX ports
  - OSA port cabling to multiple switches – part of the zEnterprise infrastructure

IEDN implementation scenario 1: New multi-tier workload

Single IEDN VLAN, new zEnterprise application workload, z/OS as the gateway to the external network

Hosts/routers on the external network can access the DVIPA addresses as long as they listen to dynamic route updates (as normal)



Hosts on the IEDN can access the dynamic VIPA addresses without using indirect routing, only when they come out of the same IP subnet as the IEDN hosts are attached to

New IEDN workload considerations

- Assume in this scenario, a single VLAN on the IEDN
- Assume also, the IPv4 subnet assigned to that IEDN VLAN is 10.1.1.0/24
- As long as the z/OS DVIPA addresses are assigned out of the same IPv4 subnet, the IEDN hosts can access them
  - OSX interface recovery and DVIPA movement addressed via ARP updates – as discussed earlier
- If the IEDN hosts need a route to the external network, a VIPADEFINED DVIPA can be used as the default router from the IEDN hosts
  - The z/OS default router responsibility will move to a backup z/OS if primary z/OS fails – retaining access to he external network for the IEDN hosts
- z/OS should be defined with MULTIPATH PerConnection to load-balance outbound traffic from z/OS

OMPROUTE and the IEDN

- If you want to prevent external traffic from being routed to the IEDN VLAN, then do one of the following so that OMPROUTE does not advertise the intraensemble subnet:
  - Define the OSX interface to OMPROUTE using an INTERFACE statement or IPV6_INTERFACE statement, and do not enable the IMPORT_DIRECT_ROUTES function of AS boundary routing.
  - Do not define the OSX interface to OMPROUTE, and ensure that GLOBAL_OPTIONS IGNORE_UNDEFINED_INTERFACES is configured to OMPROUTE..
- If you want to allow external traffic to be routed to the IEDN VLAN, then define the OSX interface to OMPROUTE as an OSPF_INTERFACE or IPV6_OSPF_INTERFACE, and code a nonzero value for the ROUTER_PRIORITY parameter on the interface.
  - As long as no other hosts on that OSX VLAN have coded their interfaces as OSPF interfaces, then OMPROUTE advertises the subnet (or IPv6 prefixes) of the intraensemble data network into the OSPF network.
  - This advertisement makes all addresses that fall into the intraensemble subnet (or IPv6 prefixes) reachable using OSPF.
- **Tip**: These definitions apply per interface, so you could implement advertising on one VLAN while not advertising on a different VLAN attached to the same z/OS router.
- **Remember**: z/OS will never route traffic between IEDN VLANs

IEDN implementation scenario 2: Extending existing Sysplex workload with IEDN workload

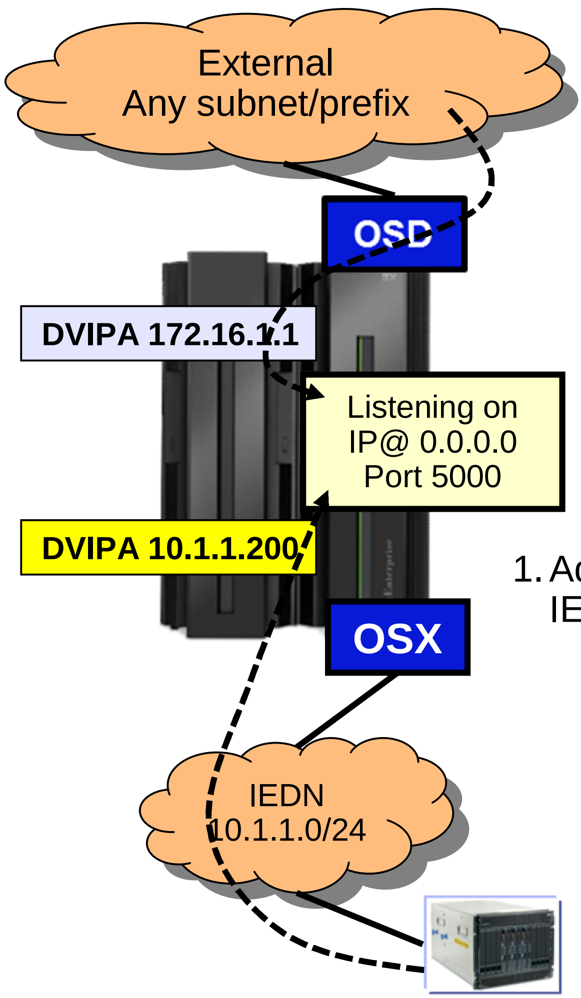
Considerations for Extending existing Sysplex workload with IEDN workload

- DVIPA addresses that must be accessed from an IEDN VLAN, should be assigned out of the same subnet as the IEDN VLAN itself.
- If you are adding IEDN access to an existing z/OS Sysplex that already use VIPA addresses for application access and you need to access those same applications from the IEDN, you have a few choices:
  1. Add static route definitions to IEDN nodes (a single default route may be enough)
     - Will work, but may result in non-optimal routing paths
  2. Renumber the affected DVIPAs to use IEDN VLAN addresses
     - Everything looks like it did in the previous scenario
     - Update your name servers with the new addresses
     - Hope (!) that everyone in the external network does use a name server when accessing those applications
  3. If the applications use generic IP address binding (0.0.0.0), you can assign new additional VIPA addresses to be used when accessing the applications from the IEDN
     - The TN3270 server may be accessed through both an existing VIPA address and a new VIPA address
  4. If the application uses a bind-specific DVIPA address, and the application supports being started on multiple addresses, define a new “instance” of the listener running on that new DVIPA address
     - This will work for many bind-specific applications, such as CICS Sockets Domain, CICS Sockets, IMS Connect, FTP servers, TN3270 servers, etc.
     - When using DB2 in data sharing mode you can define a common alias for the existing DB2 members and associate that with unique IEDN DVIPAs and unique common ports (requires DB2 V10)
- Always remember to check your networking policy rules when adding new IP addresses.

IBM

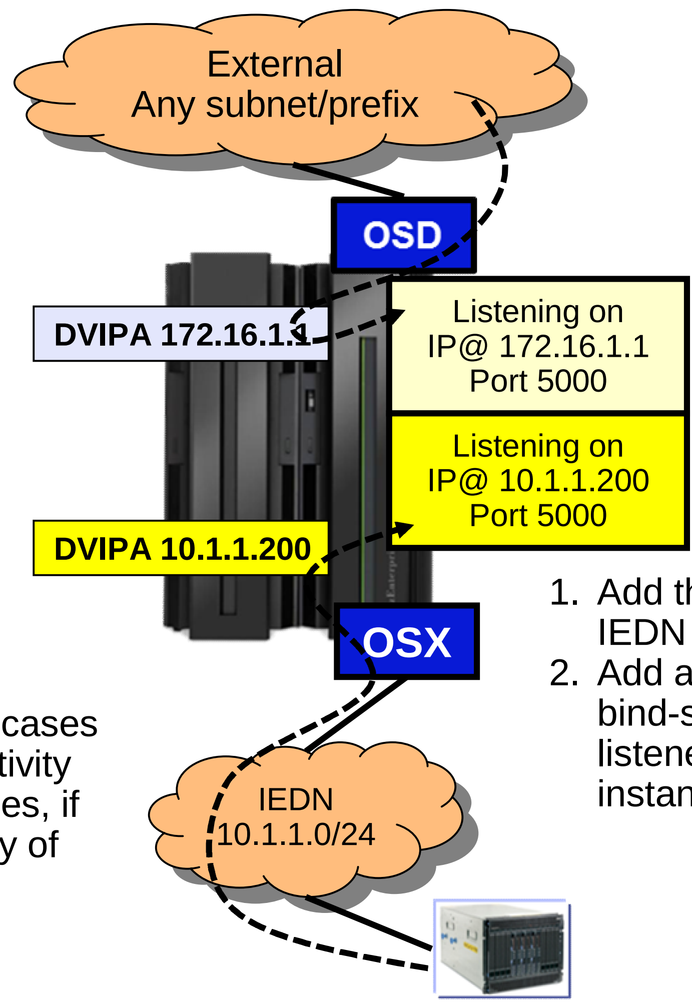
Adding IEDN access to existing z/OS servers

Generic servers



1. Add the new IEDN DVIPA

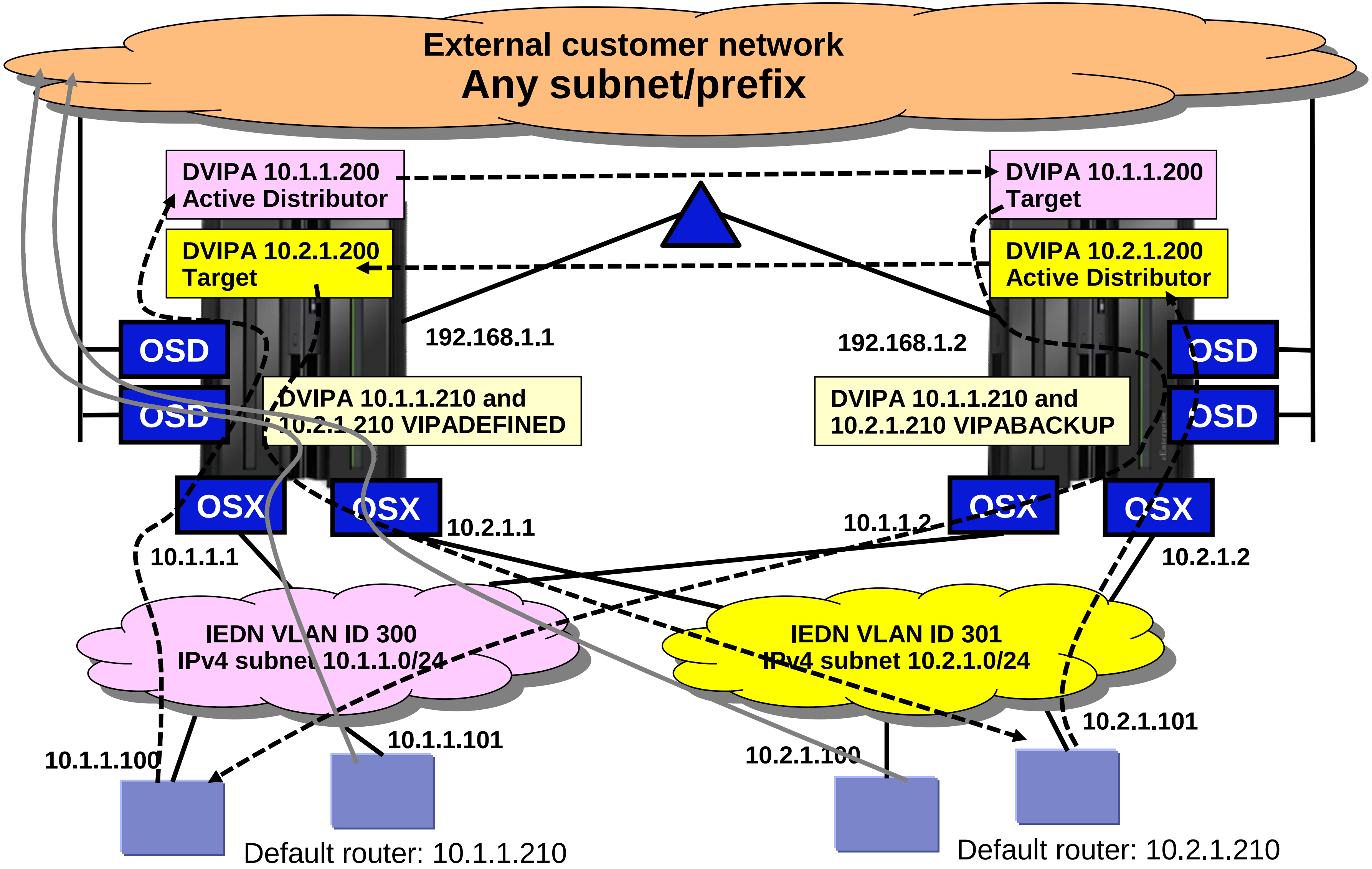
Bind-specific servers



1. Add the new IEDN DVIPA
2. Add a new bind-specific listener instance

Remember in both cases to update connectivity rules in your policies, if you are using any of those

IEDN with multiple VLANs



Considerations for multiple VLAN IEDN access

- Basic rule is that IEDN hosts can access DVIPAs in their own IEDN VLAN subnet/prefix
  - Exception is if a z/OS system is used as the default router, in which case the default router z/OS system may know how to reach the DVIPA from another IEDN VLAN
    - This may result in non optimal routing: route to the default router, forward to the distributing z/OS system, connection forwarding to the target z/OS system, which may or may not have direct access to the source IEDN VLAN (in which case another routing hop via another z/OS system will be needed on the way back)
      - But – it will work
    - You can control at a very detailed level what amount of routing z/OS is to perform in these cases through IPSec filter rules on z/OS
- Each IEDN VLAN must have a default router DVIPA defined on z/OS - if z/OS is used as the gateway to the external network
- **Remember**: z/OS will never route traffic between IEDN VLANs

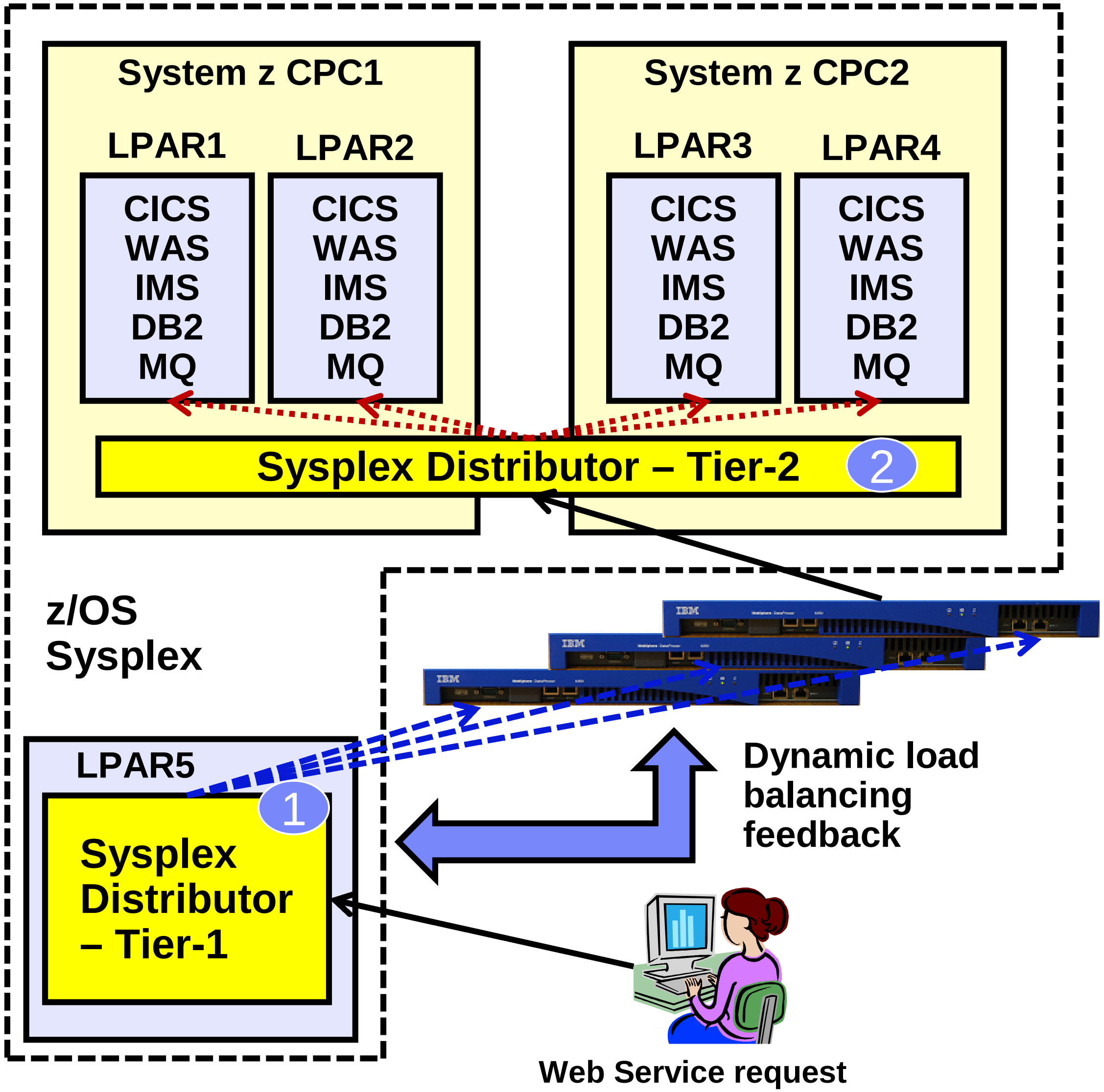
zEnterprise networking – z/OS IEDN network design and implementation (Part 2)

IEDN implementation scenario 3: DataPower and Sysplex Distributor

Introducing the WebSphere DataPower XI50z for zEnterprise

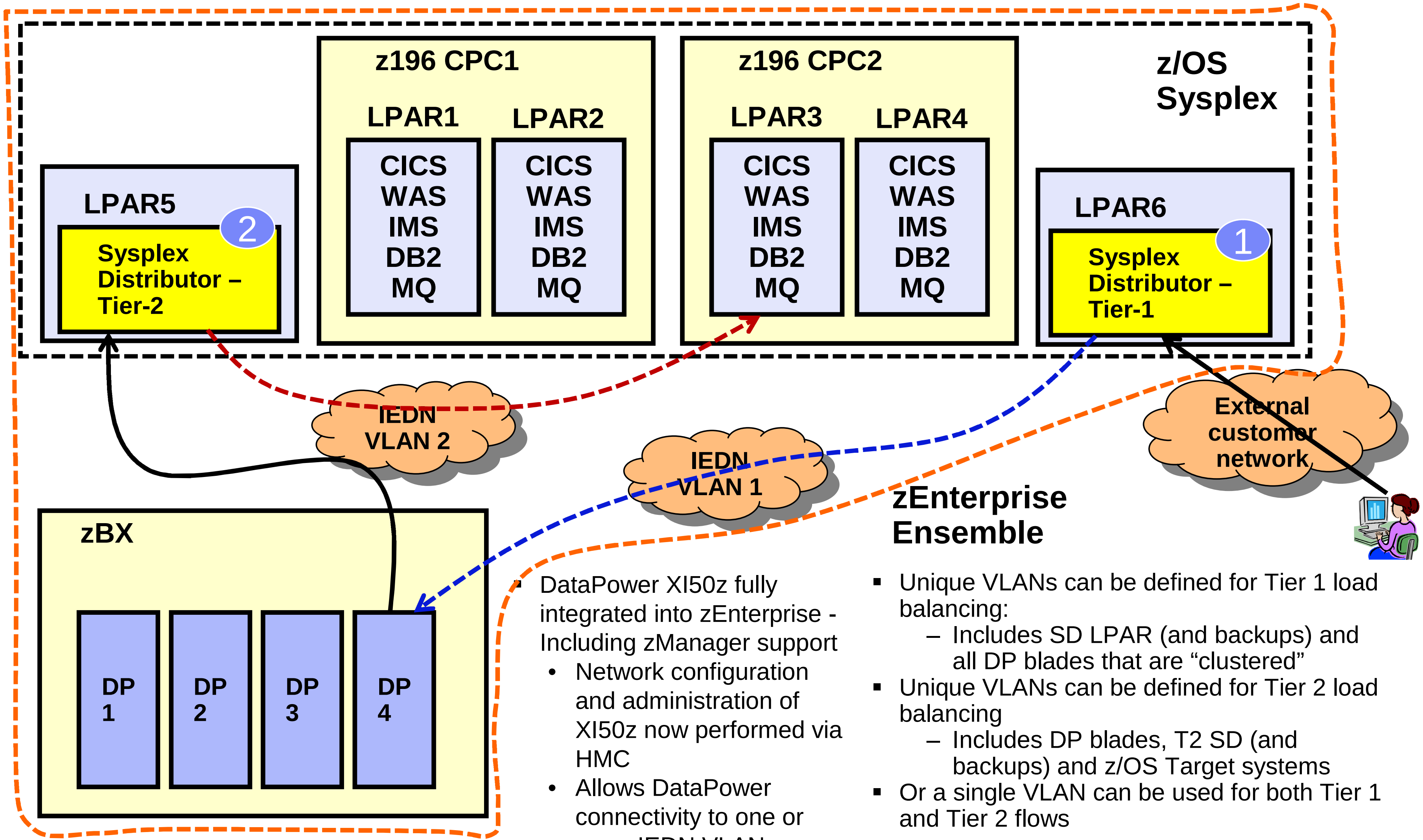
- XI50z features optimized in a dense, high compute IBM zEnterprise BladeCenter Extension (zBX) form-factor
- Supports all ESB, Security, and Integration capabilities of DataPower XI50 v3.8.1
- **Purpose-built** Integration Appliance
  - *Sysplex,* CICS, IMS, DB2, SAF, RACF integration
- **Highest capacity** DataPower appliance for SOA workloads optimized for zEnterprise environments
- Tightly **integrated** with zEnterprise
  - Unified hardware and firmware management through the Hardware Management Console (HMC)
  - Inherits serviceability, monitoring and reporting capabilities of zEnterprise

Sysplex Distributor support for DataPower



- Introduced in z/OS V1R11 Communications Server
  - DataPower Support in Firmware 3.8.1
- Allows Sysplex Distributor to load balance connections to a cluster of DataPower appliances that "front-end" a z/OS Sysplex environment (Tier 1)
  - Complements Sysplex Distributor support for back-end workflows (DataPower to z/OS – Tier 2)
- Sysplex Distributor and DataPower communicate over a control connection
  - Allows SD to have awareness of state and utilization levels of each DataPower instance
  - Facilitates TCP connection management and use of GRE to preserve client's IP address visibility to DataPower

WebSphere DataPower XI50z – Sysplex Distributor use case



- DataPower XI50z fully integrated into zEnterprise - Including zManager support
  - Network configuration and administration of XI50z now performed via HMC
  - Allows DataPower connectivity to one or more IEDN VLANs

- Unique VLANs can be defined for Tier 1 load balancing:
  - Includes SD LPAR (and backups) and all DP blades that are "clustered"
- Unique VLANs can be defined for Tier 2 load balancing
  - Includes DP blades, T2 SD (and backups) and z/OS Target systems
- Or a single VLAN can be used for both Tier 1 and Tier 2 flows

zEnterprise networking – z/OS IEDN network design and implementation (Part 2)

IEDN enabled HiperSockets Configuration Considerations

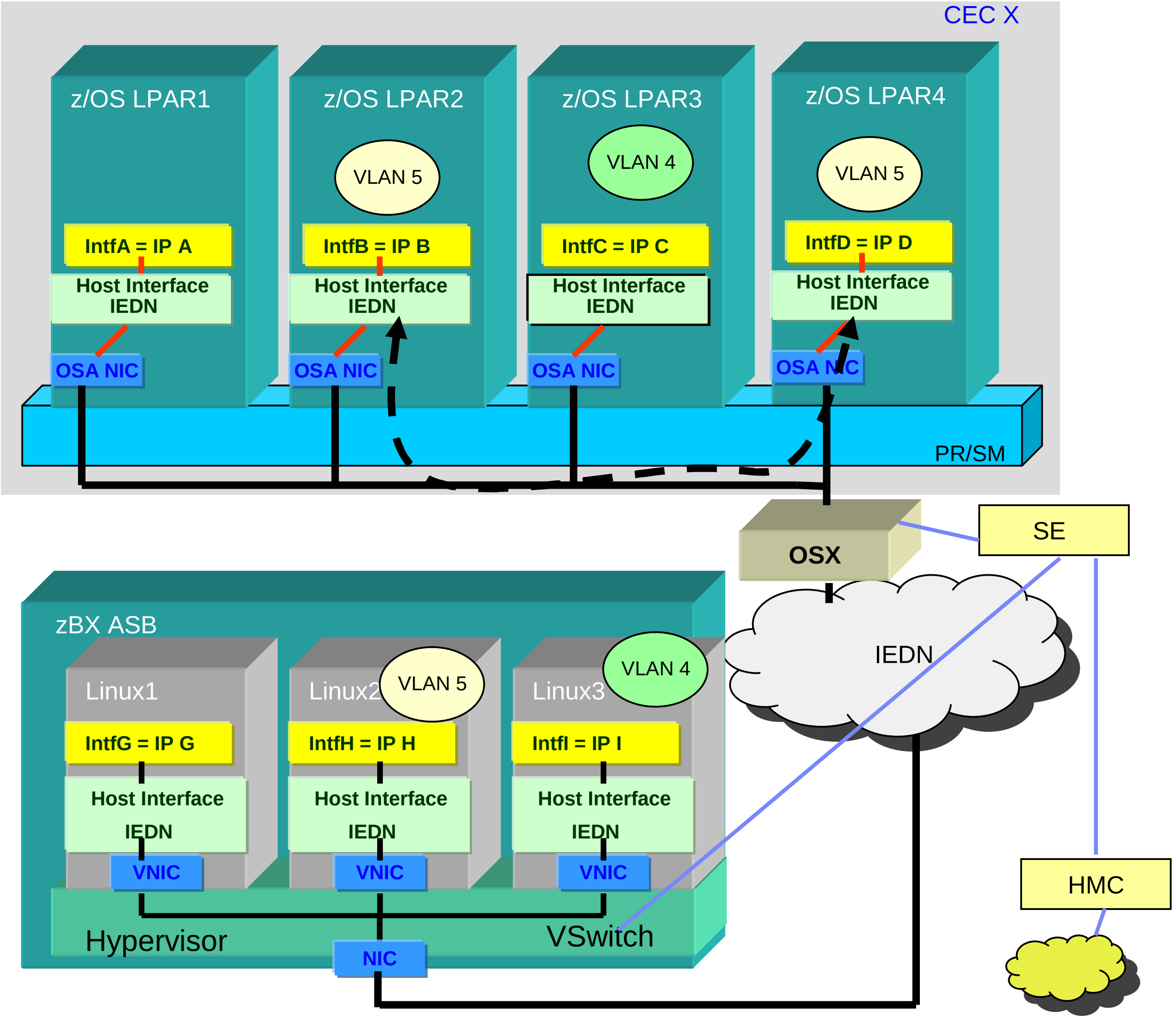
1. HiperSockets – Internal CPC network that provides high performance (low latency) communications for LP to LP communications

2. IEDN – Intra-Ensemble Data Network that provides secure and high performance (10GbE) communications within the Ensemble

… the challenge… bring the value propositions of both technologies together (**objective = create a “single internal network**”).

zEnterprise IEDN without Hipersockets

.... Intra Ensemble Data Network with platform managed virtualization, isolation and access controls

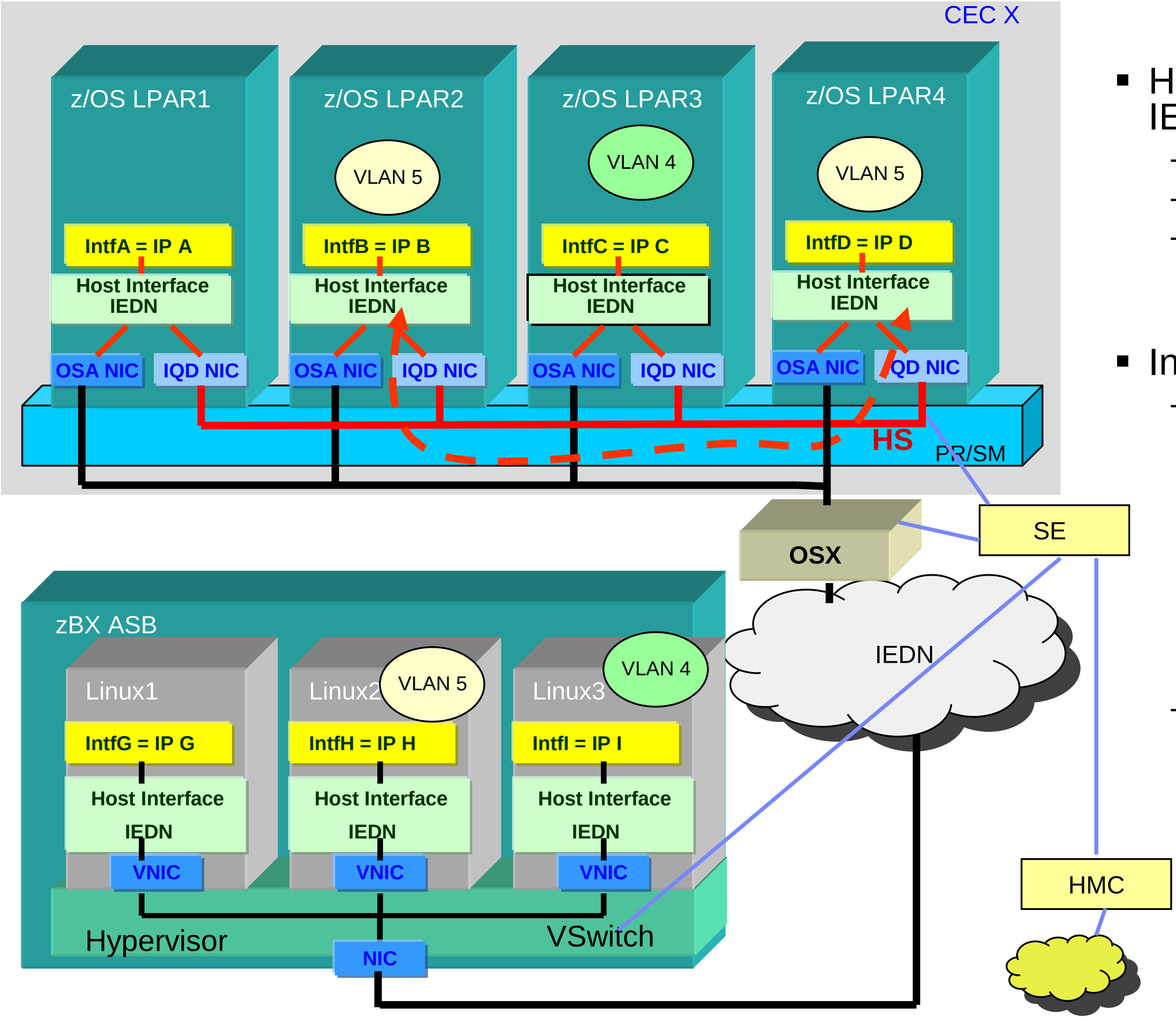


HiperSockets is another type of System z internal network that is a System z differentiator!

…yet HS is missing from the IEDN … in order to exploit HS it requires explicit and separate network config (IP address, IP route, OS config etc.)

IEDN enabled HiperSockets - IQDX

.... Intra Ensemble Data Network with platform managed virtualization, isolation and access controls

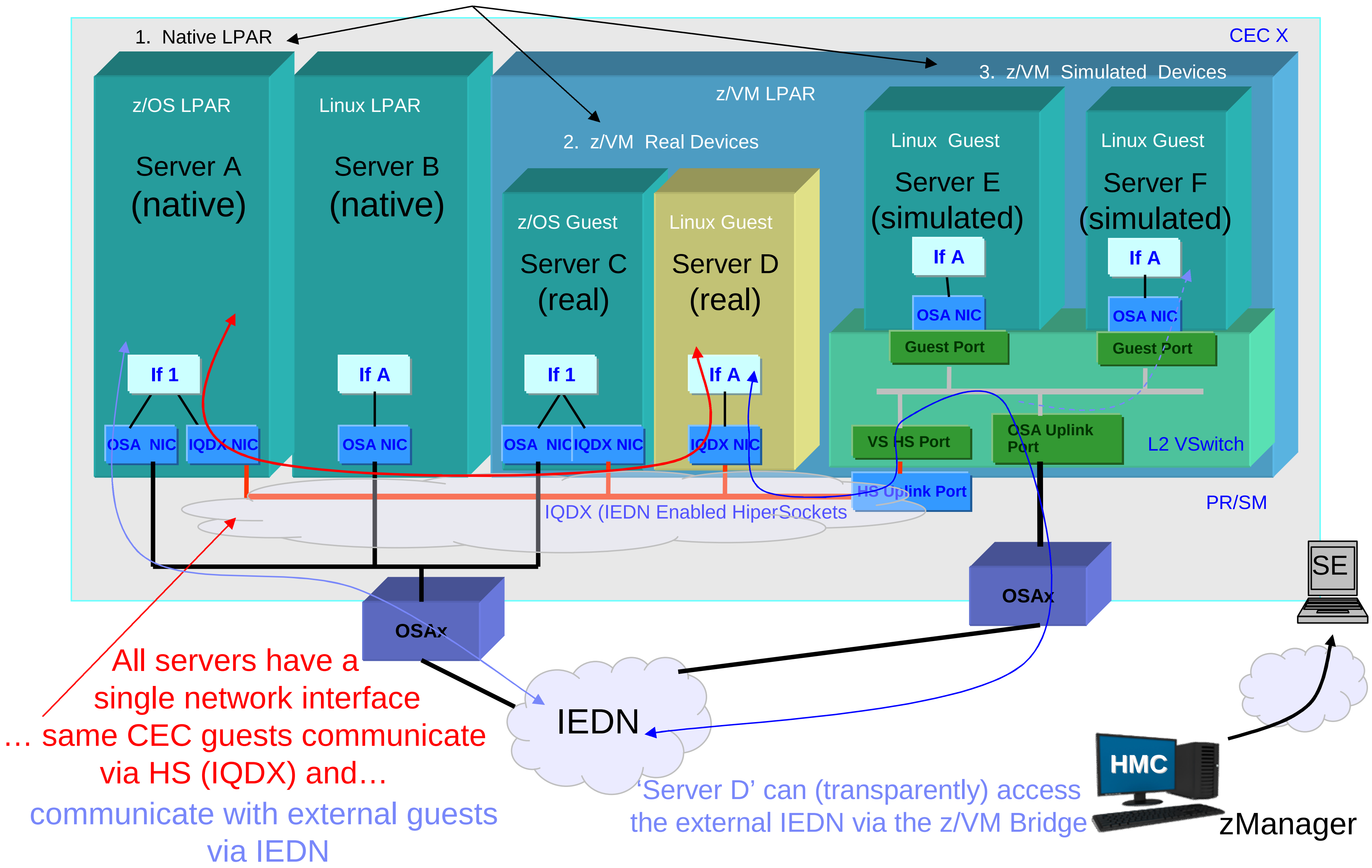


- HiperSockets becomes part of the IEDN
  - z/OS support in V1R13
  - zVM support in zVM 6.2
  - zEnterprirse support required as well (see announcement)

- In a transparent manner
  - The virtual servers present a single IP address (their IEDN address) for both internal (HiperSockets) and external (IEDN) access
    - No IP topology changes or routing changes required
    - The optimal path is selected automatically without requiring unique routing configuration
  - Also enables relocation of System z virtual servers across z CECs without reconfiguration
    - Same IP address used
    - Current HiperSockets IP topology is CEC specific
      - Moving to another CEC requires IP address and routing changes.

For more information on recent announcement:
http://www-01.ibm.com/common/ssi/rep_ca/6/897/ENUS112-026/ENUS112-026.PDF

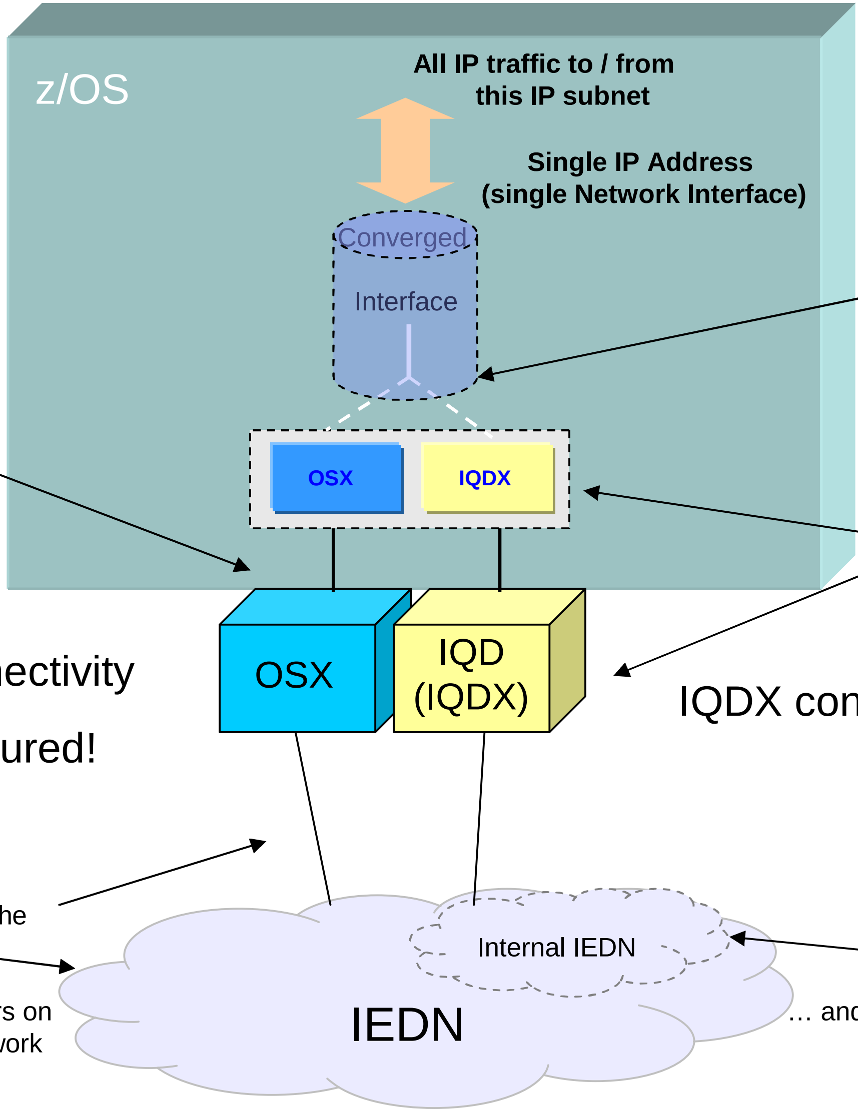
Complete IEDN Enabled HS zEnterprise Solution (provided for all 3 types of guest virtual machines)



System z/OS “Converged IQDX Link” - Concepts

The z/OS administrator
only configures / manages
the OSX interface

The zEnterprise Ensemble
network administrator
only configures VLANs for OSX



Communications Server transparently
splits and converges
network traffic to this interface

With **AUTOIQDX…**
the IQDX interface is
dynamically / transparently
configured / added
(“tucked” under OSX)

Only OSX connectivity
must be configured!

IQDX connectivity is transparent!

The OSX link provides access to the
“External IEDN fabric”

… and provides access to all servers on
both the Internal and External Network

The IQDX link provides access to the
“Internal CPC IEDN fabric” (optimal path)

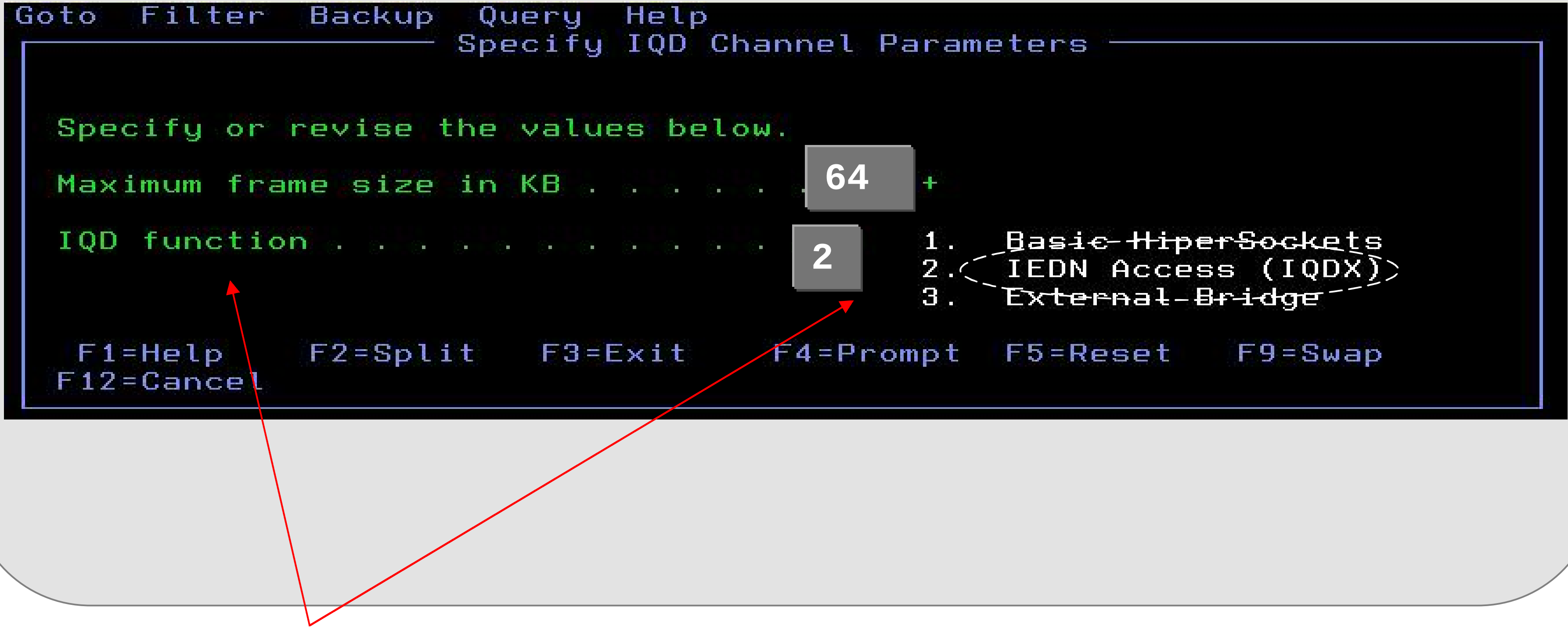
… and provides access to just the subset of servers on the
“Internal (IQDX portion of the) IEDN”

IEDN Enabled HS - Key Single Network Usability Points

Enablement …. IQDX applies to a single IQD CHID (per CPC)

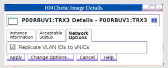
1. System I/O Configuration
   - Select (reuse or configure) an IQD CHID and add new IQDX (HCD) Channel Parameter
     - Note: existing IQD HCD parameters are required / unchanged (LPs, devices, etc.)
2. HMC (zManager) – Replicate VLAN IDs (default setting)
   - no HS network config tasks are required (IQDX VLANs are inherited from OSX) (users will have an option at the HMC (NVM) to manually configure HS)
3. z/OS:
   - Configure IEDN access via OSX (OSA interface)
   - IQDX is dynamically / transparently configured when OSX is configured and with GLOBALCONFIG AUTOIQDX parameter
4. Linux - IQDX is transparent (requires z/VM bridge solution):
   - OS – configure IEDN access via IQD (IQDX via existing HS QETH interface) (no OSA configuration – single HS interfaces provides access to the entire IEDN)
5. z/VM – Configure z/VM VSwitch Bridge Support (zVM Bridge provides external IEDN access for QEBSM guests (Linux only))

HCD Enablement – IQD Function (Option 2)



Select Option 2 to enable IEDN HS (IQDX) – valid for only one IQD CHPID per CPC

HMC (zManager) Enablement – IQDX VLANs (Replicate VLANs)



Select Replicate VLAN ID to vNICs (“Replicate” is the default setting)

z/OS Enablement – AutoIQDX (Defaults to AutoIQDX “Enabled”)

```
>>-GLOBALCONFig-------------------------------------------->

   >---------------------------------------------------------------><
     |                            .-ALLTRAFFIC--------|           |
     | .-AUTOIQDX-----+-------------------------+---------|
     | |                          '-NOLARGEDATA----|          |
     +---------------------------------------------------------------|
     | |-NOAUTOIQDX----------------------------------------|
```

“ALLTRAFFIC” Option controls large messages (streaming) over HS (NOLARGEDATA directs TCP/IP to send large data external via OSA)

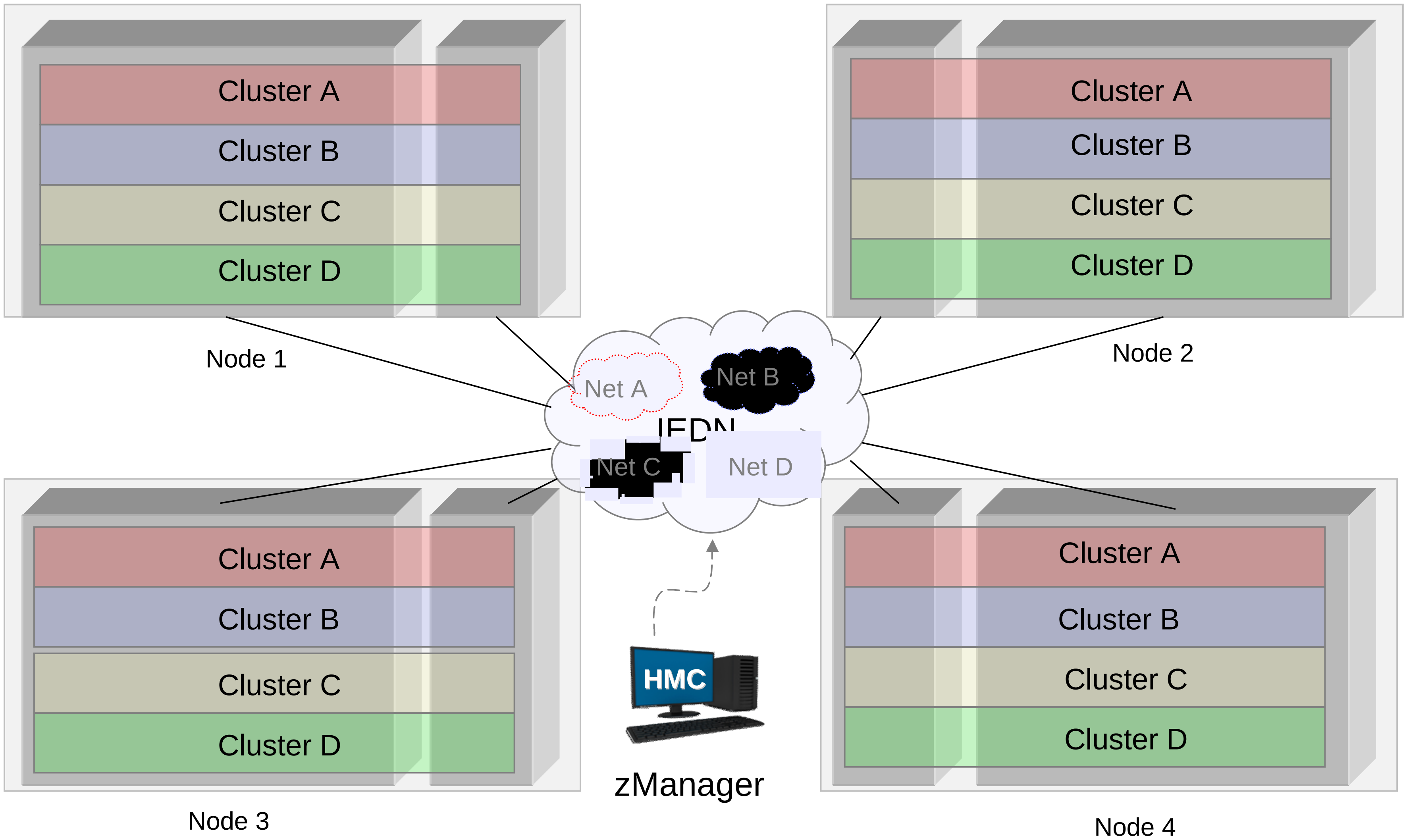
```
Global Configuration Information:
  TcpIpStats: Yes  ECSALimit: 2096128K  PoolLimit: 2096128K
  MlsChkTerm: No   XCFGRPID:  11        IQDVLANID: 27
  SysplexWLMPoll: 060  MaxRecs:   100
  ExplicitBindPortRange:  05000-06023   IQDMultiWrite: Yes
  AutoIQDX: AllTraffic
  WLMPriorityQ: Yes
```

Sample z/OS Display of OSX (with IQDX) – NETSTAT DEVLINKS

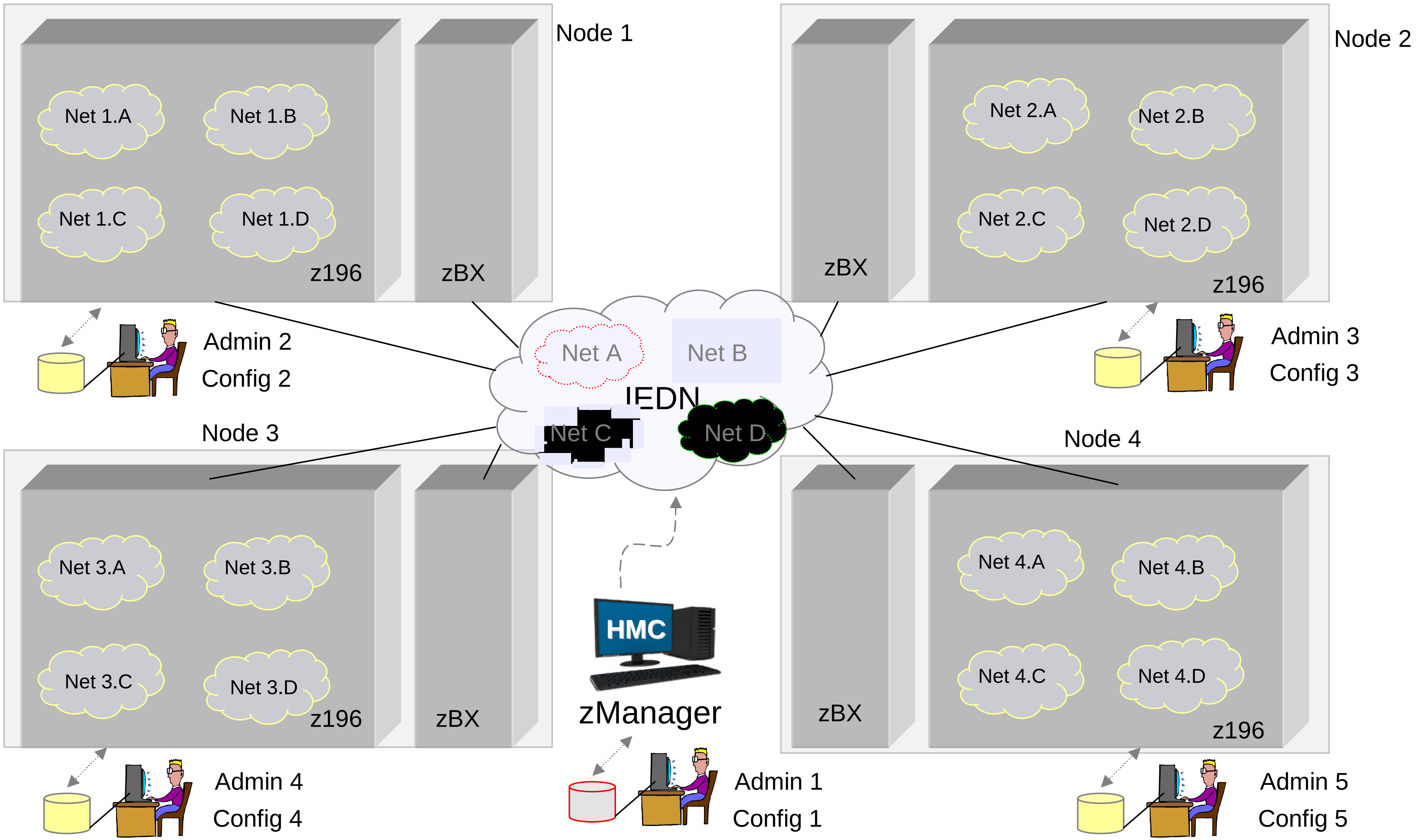
```
IntfName: OSXC9INT1         IntfType: IPAQENET  IntfStatus: Ready
    PortName: IUTXP0C9  Datapath: 0E56     DatapathStatus: Ready
    CHPIDType: OSX      CHPID: C9
    QueSize: 0    Speed: 0000001000
    VMACAddr: 620001AA0E56   VMACOrigin: OSA    VMACRouter: All

  Interface Statistics:
    BytesIn                                          = 0
    Inbound Packets                                  = 0
    Inbound Packets In Error                         = 0
    Inbound Packets Discarded                        = 0
    Inbound Packets With No Protocol                 = 0
    BytesOut                                         = 688
    Outbound Packets                                 = 7
    Outbound Packets In Error                        = 0
    Outbound Packets Discarded                       = 0
  Associated IQDX interface: EZAIQXC9  IQDX Status: Ready
    BytesIn                                          = 0
    Inbound Packets                                  = 0
    BytesOut                                         = 0
    Outbound Packets                                 = 0
```

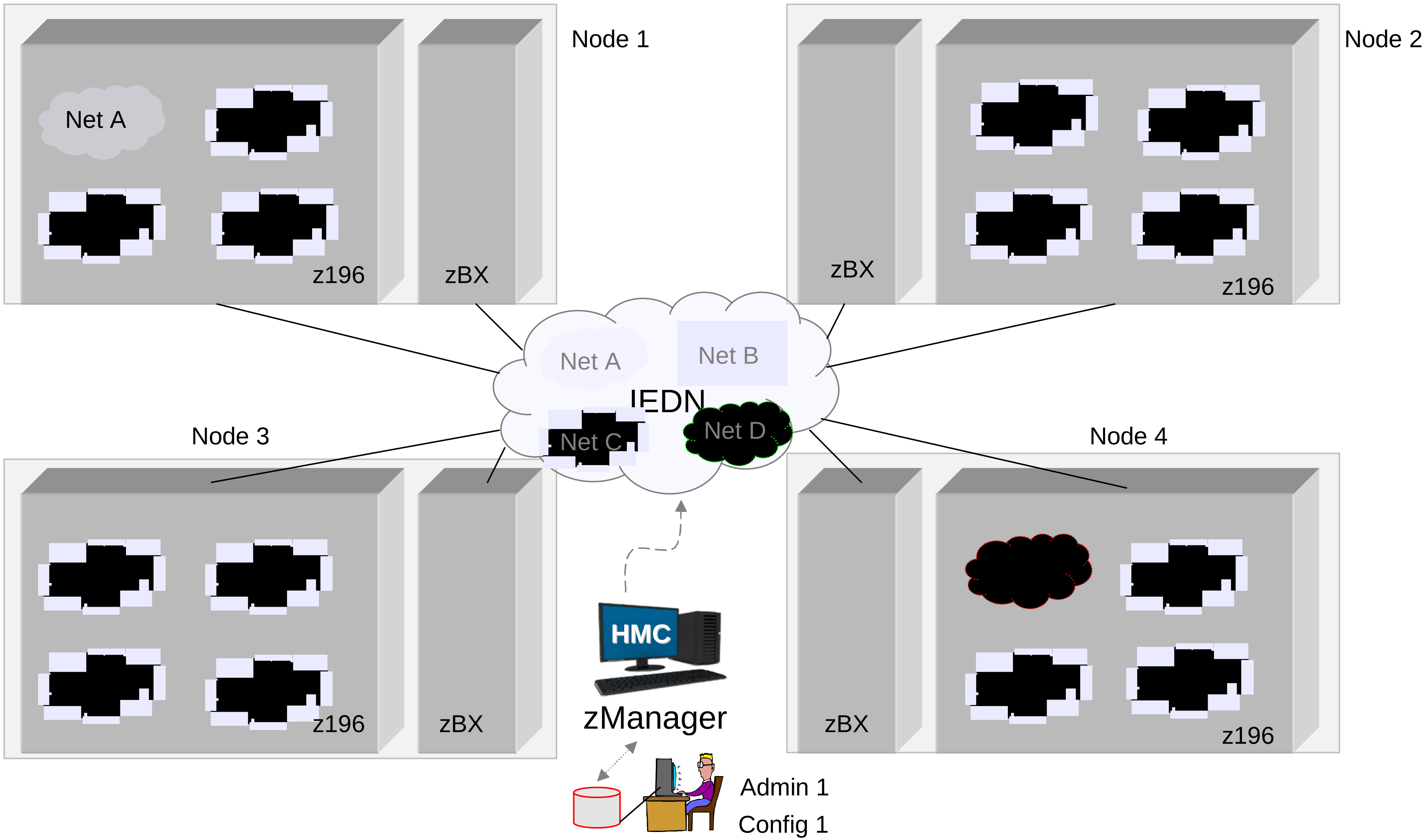
zEnterprise Ensemble (4 unique clusters (groups of virtual servers) – provisioned on all 4 CPCs each cluster exploits a unique virtual network via a single administrator)



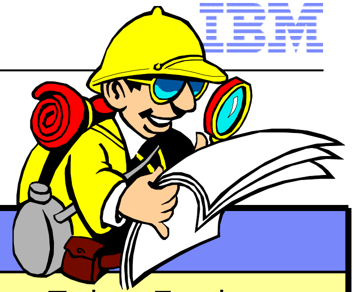
zEnterprise Ensemble with IQD
(… now the 4 clusters requires 20 networks and 5 administrators)



zEnterprise Ensemble with IQDX
(4 clusters reduced to 4 networks with a single administrator)



For more information



URL	Content
http://www.twitter.com/IBM_Commserver	IBM z/OS Communications Server Twitter Feed
http://www.facebook.com/IBMCommserver	IBM z/OS Communications Server Facebook Page
https://www.ibm.com/developerworks/mydeveloperworks/blogs/IBMCommserver/?lang=en	IBM z/OS Communications Server Blog
http://www.ibm.com/systems/z/	IBM System z in general
http://www.ibm.com/systems/z/hardware/networking/	IBM Mainframe System z networking
http://www.ibm.com/software/network/commserver/	IBM Software Communications Server products
http://www.ibm.com/software/network/commserver/zos/	IBM z/OS Communications Server
http://www.redbooks.ibm.com	ITSO Redbooks
http://www.ibm.com/software/network/commserver/zos/support/	IBM z/OS Communications Server technical Support – including TechNotes from service
http://www.ibm.com/support/techdocs/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
http://www.ibm.com/developerworks/rfe/?PROD_ID=498	RFE Community for z/OS Communications Server
https://www.ibm.com/developerworks/rfe/execute?use_case=tutorials	RFE Community Tutorials

For pleasant reading ….

Please fill out your session evaluation

- zEnterprise System - z/OS IEDN network design and implementation
- Session # 11335
- QR Code:

IBM Comm Server

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

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

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

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