IBM® ATS zEnterprise™ System

Get Ready for a New Dimension in Computing.
System z Technical Education

IBM zEnterprise System

Network Virtualization, Management, and Security

Gwendolyn J. Dente, IT Specialist ATS
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.
IBM® ATS zEnterprise™ System Technical Roadshow

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.


Notes:

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

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.

Call Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefore.

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 Trade Association and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association, Inc., Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Core, Intel SpeedStep, Intel Xeon, 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 Linux 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.

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

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 products 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.

© 2011 IBM Corporation
Abstract

You've heard a lot about the IBM zEnterprise™ System. The new machines are faster, more powerful and more energy efficient. But the most significant change is that other kinds of computers can now be "plugged into" the mainframe to create an "Ensemble Network" where security exposures are minimized and the data center can be managed as if it were a single computer. Many questions about speeds, feeds, feature codes, operating system levels have been answered, but many more questions have been raised about network design and network security. Attend the sessions in a two-part series to hear the answers to questions about Ensemble networking: questions on the underlying architecture, on the routing and security structures, and on the software definitions.

The first session in the series, "IBM zEnterprise System Network Virtualization, Management, and Security (Part 1: Overview)," presents a high-level overview of the networking topics surrounding the new architecture. Part 1 is suitable for both an executive and a technical audience with both architects and implementers represented.

The second session, "IBM zEnterprise System Network Virtualization, Management, and Security (Part 2: Detail)," presents a more detailed view of the underlying architecture, its routing and security structures, and some of its software definitions. Part 2 is suitable for a technical audience that wants to understand more about the design, positioning, and implementation of the new architecture.

Both documents are available at: w3.ibm.com/support/techdocs
Acknowledgments

From IBM z/OS® Communications Server Development:
  Kim Bailey
  Mike Fox
  Jeff Haggar
  Bebe Israel
  Gary McAfee
  Jerry Stevens

From IBM z/VM® Development:
  Alan Altmark
  Angelo Macchiano
  Rick Tarcza
  Romney White
  Richard G. Young

From IBM International Technical Support Organization (ITSO)
  Mike Ebbers
  Parwez Hamid
  Karl-Erik Stenfors
  William G. White

From IBM Advanced Technical Skills (ATS)
  Mary Astley
  Harv Emery
  John Goodyear
  Gregory Hutchison

From IBM S&D Group, Systems Software Development
  Friedrich M Welter

Thanks to the many people and resources I found to help me understand the information I used to write this presentation.
Business processes and the applications that support them are becoming more service oriented, modular in their construction, and integrated.

The components of these services are implemented on a variety of architectures and hosted on heterogeneous IT infrastructures.

Approaches to managing these infrastructures along the lines of platform architecture boundaries cannot optimize:

- alignment of IT with business objectives;
- responsiveness to change;
- resource utilization;
- business resiliency;
- or overall cost of ownership.

Customers need a better approach: The ability to manage the IT infrastructure and Business Application as an integrated whole.

With this visual, you see what has led to the revolutionary introduction of the zEnterprise™ and the zBX with its non-z hardware blades into the System z family. We need to simplify the management of all the complex processes that comprise an IT installation by integrating and centralizing the functions that are needed to support an enterprise. At the same time, we need to extend the superior security, high availability, and performance features of IBM® System z® to business processes that run on distributed platforms.
Customers need better approach: The ability to manage the IT infrastructure and Business Application as an integrated whole.

Reduce the scope of security vulnerability in the network: many hops collapsed to fewer hops and possibly only one hop

With this visual you see how you might take new workloads on distributed systems and incorporate them into a zEnterprise Ensemble. We need to simplify the management of all the complex processes that comprise an IT installation by integrating and centralizing the functions that are needed to support an enterprise. At the same time, we need to extend the superior security, high availability, and performance features of IBM® System z® to business processes that run on distributed platforms.
The zEnterprise™ System addresses all three client imperatives for achieving a Smarter Planet. Each of the initiatives is reflected in the architecture of the Hybrid Solution. Pervasive themes in the zEnterprise System design architecture are: Virtualization, Centralization, and Redundancy, which lead to Simplification, enhanced Performance, enhanced Availability, enhanced Security, and enhanced Management.
zEnterprise Value

1. Network Simplification (“Network in a Box”)
   - Single physical network and IBM zEnterprise BladeCenter® Extension (zBX) “package” (physical integration)
   - “Ensemble” and Clustering of Heterogeneous Resources & Capabilities
     - Centralization of Appliances, Distributed Systems at zEnterprise
   - Central point of Management for heterogenous platforms (HMC/SE)
   - Reduced network path length; reduced number of hops
   - Co-location of Business Processes and Business Data
     - Reduce path length and latency issues
   - “Fit for Purpose” (select the optimal platform for the workload type)

2. Secure communications
   - Physical security (internal network equipment)
   - Logical security (controlled access through centralized definition of the Hypervisor functions at Unified Resource Manager)
   - Network Virtualization and Isolation within Internal Networks

3. High Availability
   - Redundant Network Hardware
   - Logical failover

4. Unique IBM System z® QoS and Performance
   - Exploit centralized security of RACF®
   - Improved throughput (see 1)

5. Increased Opportunity for Collaboration of Technical Talent
   - Broadening the base of System z Skills
   - Make System z relevant to a new IT Generation

The IBM zEnterprise™ System is a first-of-a-kind workload-optimized multiplatform (or multi-architecture) computing environment that spans (and tightly integrates) mainframe and distributed technologies. This system (of systems) consists of an IBM zEnterprise™ 196 (z196), the IBM® zEnterprise™ BladeCenter® Extension (zBX) Model 002, and the IBM zEnterprise™ Unified Resource Manager (“Unified Resource Manager”). The z196 is designed with improved scalability, performance, security, resiliency and availability. The Unified Resource Manager, working with the z196, the zBX infrastructure, and the attached blades, can help to deliver end-to-end virtualization and management providing the ability to align the technology deployment environment according to individual workload requirements.
This visual summarizes in a different fashion the same points made on the previous pages.

The IBM zEnterprise™ System is a first-of-a-kind workload-optimized multiplatform (or multi-architecture) computing environment that spans (and tightly integrates) mainframe and distributed technologies. This system (of systems) consists of an IBM zEnterprise™ 196 (z196), the optional IBM® zEnterprise™ BladeCenter® Extension (zBX) Model 002, and the IBM zEnterprise™ Unified Resource Manager (“Unified Resource Manager” firmware). The z196 is designed with improved scalability, performance, security, resiliency and availability. The 40% improvement is the overall, average increase in performance for workloads running on z196 compared to running on z10. The Unified Resource Manager, working with the z196, the zBX infrastructure, and the attached blades, can help to deliver end-to-end virtualization and management providing the ability to align the technology deployment environment according to individual workload requirements. The Unified Resource Manager also provides the differentiator between a stand-alone Blade Center attached via a 10-gigabit network over a single networking hop to a System z and the zEnterprise System solution. Whereas both the stand-alone configuration just described and the zEnterprise System with a zBX provide similar enhancements in terms of network performance -- to a certain extent -- in Security, the added value of zEnterprise Ensemble networking lies in the centralization of administration and security, and the simplified administrative Graphical User Interface, which can drive down training costs, security exposures, and migration timeframes.

As a result, an IT organization can realize tangible and intangible benefits from a zEnterprise deployment.
AGENDA

- Networking Architecture
  - IBM System Analytics Optimizer (ISAOPT)
  - zEnterprise and zBX Model 002

- Ensemble Overview
  - Terminology: Nodes, IBM zEnterprise Unified Resource Manager

- Networks in the Ensemble
  - The intranode management network (INMN)
  - The intraensemble data network (IEDN)
    - Hardware, Protocols
    - IP Addressing and Routing
    - Sample Definitions

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represents goals and objectives only.
AGENDA …

- Network Topologies in the Ensemble
  - Real Devices and Simulated Devices
  - Sharing and Isolating the OSA Adapters
  - Connection Redundancy
  - Sysplex Distributor Considerations
  - IP Forwarding
  - Connecting with the External Customer Network
- Secure Networking with VLANs in the Ensemble
- Security Layers in the Ensemble
- What may Surprise You When Implementing an Ensemble
- Appendix: Additional Topics for Ensemble Networking
- References (Hardware & Software)

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represents goals and objectives only.
A zEnterprise (a z196) + (zBX) + (Unified Resource Manager) = a “zEnterprise System”

When IBM zEnterprise BladeCenter Extension (zBX) is attached to IBM zEnterprise 196, the zBX infrastructure works with the IBM zEnterprise 196 to support the multiplatform environment, combining mainframe and distributed technologies. As you see in the visual, the zBX can support IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1 (5697-AQT) and POWER7 blades. The IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1 (5697-AQT) is designed to deliver improved service through accelerated and accurate business insight. This workload-optimized, appliance-like add-on is installed in the zBX, enabling application integration with System z transaction processing, messaging, and data serving capabilities. The blades are managed as a single logical virtualized environment by IBM zEnterprise Unified Resource Manager. The initial solution for creating an integrated, centralized, high-speed, low-latency solution for processing complex DB2 queries is to integrate the IBM Smart Analytics Optimizer (ISAOPT) into the IBM® zEnterprise™ BladeCenter® Extension (zBX) with or without Power Blades (“PWRBLADE”) in a second chassis of the zBX. The DataPower Appliance (XI50Z) was also announced for support within the zBX on February 15, 2011. There is a Statement of Direction (SOD) for System x Blades as blades in the Blade Center Chassis of the zBX.

BladeCenter Chassis. A modular chassis that can contain multiple blades, allowing the individual blades to share resources such as the management, switch, power, and blower modules. On zEnterprise, an IBM blade, installed in the zBX, can be used to host applications and enable application integration with System z transaction processing, messaging and data serving capabilities. Each zBX can have up to four BladeCenter Racks and each rack can contain up to two blade center chassis, with each containing up to 14 blades. There are 2 types of BladeCenter Chassis that can reside in the zBX: the zBX can contain optimizers (also known as Accelerators) and IBM Blades, with a maximum of 2 Blade Center chassis per rack.

IBM Smart Analytics Optimizer. An optimizer that processes certain types of Data Warehouse queries for DB2® for z/OS V1.1 (697-AQT). It is supported in the zBX, providing acceleration to System z DB2 workloads on a z10. In a zEnterprise and with a zBX Model 002, it is enabled and managed, as part of the ensemble, by the Unified Resource Manager.

Management TOR switch (Mgmt TOR switch). This Top-of-Rack 1000BASE-T Ethernet switch operating at 1 Gbps is available in the zBX Model 002. With zBX Model 002, this switch provides a private connection to the customer’s z196 SE for hardware management tasks, but is also used to provide connectivity to the ensemble intranode management network (INMN). Hardware management functions -- such as physical inventory discovery, hardware operational controls, service updates, and repair and verify tasks -- are available on the z196 with the Model 002, but connectivity through the management TOR switch is also used to provide additional management functions through the Unified Resource Manager across the INMN. There are two 1000BASE-T Ethernet switches per zBX Model 002. The TORs reside in the first rack of a zBX. (Up to 4 racks comprise a zBX.)

Data TOR switch. This is a Top-of-Rack switch for a 10 Gigabit data network. The TORs reside in the first rack of a zBX. (Up to 4 racks comprise a zBX.)

NOTE on Optimization: A z10 can access the Solutions, but can’t be part of the managed ensemble.
In this visual you see that ISAOPT accelerator or optimizer blades (both “Coordinator blades” and “Worker blades”) are installed in a zBX Model 002 that is connected with a IBM® zEnterprise™ z196. When the zEnterprise participates in the networking in this fashion, we speak of an “Ensemble” topology. An “Ensemble” requires additional firmware instructions only available starting with the zEnterprise and the zBX Model 002 in order to provide enhanced management capabilities. The Ensemble resources are managed via the Hardware management Console (HMC) with the IBM zEnterprise™ Unified Resource Manager (“Unified Resource Manager”). There must be a Primary HMC and an Alternate HMC which takes over if the Primary HMC fails.

Per Blade Center Chassis only one type of Blade is supported. (For example, you cannot mix ISAOPT optimizers with Power7 blades in the same chassis.)

When an IBM zEnterprise BladeCenter Extension (zBX) is attached to a zEnterprise 196, the zBX infrastructure works with the z platforms to support the multiplatform environment, combining mainframe and distributed technologies. The IBM Smart Analytics Optimizer for DB2 for z/OS, V1.1 (5697-AQT) is designed to deliver improved service through accelerated and accurate business insight. This workload-optimized, appliance-like add-on is installed in the zBX and connects to DB2, providing transparency to all applications.

DB2 on the mainframe must be enabled for IBM System Analytics Optimizer (ISAOPT). The Data Studio application on a workstation communicates through the mainframe to the coordinator and worker blades on the zBX to retrieve the accelerated business information.

When implemented on a z196 and the zBX Model 002, these z platforms can use operational controls on the HMC to manage IBM Smart Analytics Optimizer for DB2 for z/OS V1.1 (5697-AQT). This is a heterogeneous Business Intelligence (BI) infrastructure that is a single entity with minimal risk and improved costs. The HMC also provides extended management support through the Unified Resource Manager to manage the IBM zEnterprise BladeCenter Extension (zBX) Model 002 and IBM Smart Analytics Optimizer for DB2 for z/OS V1.1 (5697-AQT) capabilities.

The HMC with Unified Resource Manager firmware manages the zBX over the intranode management network (INMN), which is connected to the Top of Rack Switches in the zBX, where the network is running at 1 Gigabit. DB2 communicates with the zBX over the intraensemble data network (IEDN) to the Electronic (Ethernet) Switch Modules (ESMs) attached to the blades in the zBX.
IBM® ATS zEnterprise™ System Technical Roadshow

IBM DataPower Appliance XI50Z for z196 and zBX Model 002

- Private Intranode management Network (INMN) (1 Gigabit)
- Private Intraensemble data network (IEDN) (10 Gigabit)
The **zEnterprise Unified Resource Manager** ("Unified Resource Manager"). Is an integrated System z management facility responsible for zEnterprise platform management. It is part of the operational controls that run on the HMC.

*Only one pair of Ensemble HMCs per an Ensemble.*

For high availability, one is a mirrored backup, and cannot be used as an Ensemble HMC, until fail-over.

*Both are attached to the Customer Managed Management Network.*

The Unified Resource Manager provides two suites of tiered functionality – Manage and Automate.

The majority of the Unified Resource Manager (z/Manager) support is actually in support of hypervisors. For example: **Definitions through the Unified Resource Manager produce:** server container creation, network device allocation and configuration for the server, VLAN assignments.
The ensemble is managed by a System z Hardware Management Console running firmware called the IBM zEnterprise™ Unified Resource Manager. The Unified Resource Manager provides energy monitoring and management, goal-oriented policy management, increased security, virtual networking, and data management for physical and logical resources of a given ensemble.

**The Primary HMC for an Ensemble includes the Unified Resource Manager which provides two suites of functions:**


Any HMC can manage up to 100 z196s. The primary HMC can perform all non-ensemble HMC functions on z196s that are not members of the ensemble (up to the 100 z196s limit which includes any ensemble members).

A primary HMC is the only HMC that can perform ensemble-related management tasks (create virtual server, manage virtual networks, create workload, ...)

Any other HMC can only perform non-ensemble tasks on a z196 that is a member of an ensemble.

A customer can have multiple ensembles - they would need a Primary and an Alternate HMC pair for each ensemble they create.

**Security in the HMC and the Unified Resource Manager:**

The HMC continues to enjoy the security controls inherent in the System z Console. Please see IBM System z Hardware Management Console Security White Paper by Kurt Schroeder (schroedk@us.ibm.com) published in Sept. 2008 and available at:


This paper explains: "The HMC Licensed Internal Code includes a full-function firewall that is used to control network access to the HMC. As previously described, by default the HMC allows for virtually no inbound network traffic. As different features of the HMC are enabled (e.g., remote access, SNMP based automation etc.) additional inbound network traffic is allowed."

In addition, the Unified Resource Manager screens of the console enjoy additional security controls, like role-based security and operations over the closed, Intranode Management network (INMN).
In this visual you see how formerly Distributed Servers can be consolidated at a zEnterprise site to form a hybrid environment in an Ensemble that can be managed as a federated whole with selected LPARs in the zEnterprise. Appliances like ISA Optimizer and DataPower can be integrated as well by housing them in the blades of the zBX. This type of consolidation not only can reduce the path length from multiple hops to reach a server in the external network to a single hop to reach the same server in a zBX, but it can also transform the data path into an internal path that can be inherently more secure than a path over the external network.

The “Virtual Server” is a “Virtual Entity” that is defined to the Ensemble. As a result, z/OS is not a virtual server but rather an Operating System inside the Virtual Server; you can think of a native LPAR as a Virtual Server. On the other hand, under zVM, the Guest Machine would be the Virtual Server and Linux or z/OS would be the Operating Systems in that Virtual Server.

Recall what you heard earlier: “The IBM zEnterprise™ System is a first-of-a-kind workload-optimized multiplatform (or multi-architecture) computing environment that spans (and tightly integrates) mainframe and distributed technologies. This system (of systems) consists of an IBM zEnterprise™ 196 (z196), the IBM® zEnterprise™ BladeCenter® Extension (zBX) Model 002, and the IBM zEnterprise™ Unified Resource Manager (Unified Resource Manager). The z196 is designed with improved scalability, performance, security, resiliency and availability. The Unified Resource Manager, working with the z196, the zBX infrastructure, and the attached blades, can help to deliver end-to-end virtualization and management providing the ability to align the technology deployment environment according to individual workload requirements.”

Notice the dotted line between one of the Blade Optimizers and the IEDN; notice the absence of a connection to the IEDN by the other Blade Optimizer. Depending on the type of blade deployed, you may or may not need to provide a connection between the optimizer and the IEDN.
Terminology

- **Node or Member:**
  - A single IBM zEnterprise z196 and any optionally attached zBX. A node can be a member of only one ensemble.
  - Managed via the HMC using the Unified Resource Manager.
  - Comprised of z196s and typically includes zBX racks.

**A Node or Member**
- **zEnterprise + 0zBX**

**A Node or Member**
- **One z196 “footprint” plus zBX with 4 Racks**

An ensemble can contain up to 8 nodes managed by the zEnterprise System Unified Resource Manager (Unified Resource Manager). A node is one z196 with an optionally attached zEnterprise BladeCenter Extension (zBX). Each zBX can have up to four Blade Center racks.

The zEnterprise Unified Resource Manager enables management of an **ensemble**; an **Ensemble** is a collection of one or more zEnterprise **nodes** in which each node comprises a zEnterprise z196 and its optional attached IBM zEnterprise BladeCenter Extension (zBX).

A z196 can be a member of only one ensemble.

If a z196 has no zBX (called a “0 zBX Ensemble”), a Direct-Connect (not wrap) cable is required for the 10 Gigabit OSA port configured as CHPID Type OSX. You can also configure an Ensemble with two nodes comprised of two z196 platforms sharing a single zBX. In this case, the loopback cable on the 10Gigabit OSA OSX port is not required.
A zEnterprise Ensemble is a collection of servers running on various architectures within one or more zEnterprise nodes and optionally within a zBX. The resources of a zEnterprise System ensemble are managed and virtualized as a single pool of resources, integrating system and workload management across the multisystem, multitier, multiarchitecture environment.

A zEnterprise ensemble is composed of up to 8 members, with up to eight z196 servers and up to 896 blades housed in up to eight zBXs, dedicated integrated networks for management and data, and the Unified Resource Manager function. With the Unified Resource Manager, the z196 provides advanced end-to-end management capabilities for the heterogeneous systems housed in the zBX.

• The Ensemble is managed via the Hardware management Console that utilizes the integrated management tools for defining and managing heterogeneous servers.

An ensemble can contain up to 8 nodes.

• The HMC functions for the Ensemble communicate via the Support Element with the components across an internal management network called the Intra-Node Management Network (INMN). The SE uses the INMN to communicate with resources within the node.

• Servers within the Ensemble communicate with each other across an internal network called the Inter-Ensemble Data Network (IEDN).

• Only every 1st of 4 zBX Racks requires a Top of Rack (TOR) switch. Up to four zBX Racks may be interconnected to use the TOR in the first zBX Rack.
An ensemble can contain up to 8 nodes managed by the zEnterprise System Unified Resource Manager (Unified Resource Manager). A node is one z196 with an optionally attached zEnterprise BladeCenter Extension (zBX). Each zBX can have up to four Blade Center chassis.

The zEnterprise Unified Resource Manager enables management of an ensemble; an Ensemble is a collection of one or more zEnterprise nodes in which each node comprises a zEnterprise z196 and its optional attached IBM zEnterprise BladeCenter Extension (zBX).

A z196 can be a member of only one ensemble.

When there are multiple nodes in an ensemble, the IEDNs are interconnected via aggregation points (also called “Masters”) in the first and last zBX. This type of configuration is often referred to as a “Master-Slave” relationship. The first zBX maintains connections from one of the two 10 Gig switches to all the other members of the ensemble; the second of the two 10 Gig TOR switches in the last zBX (by convention called zBX #8) is the aggregation point back to all the ensemble nodes. A node without a zBX connects to another node’s 10 Gt Switches, whose zBX is then aggregated as just described.

An Ensemble may encompass no Sysplex, or a single Sysplex, or multiple Sysplexes.
The HMC functions for working with the Unified Resource Manager perform many tasks that otherwise a system administrator or system programmer would have to perform. Some operating systems exploit the virtualization functions more than others.
SECURITY and Unified Resource Manager: Unified Resource Manager will orchestrate various forms of platform management and virtualization by interacting with various elements of platform firmware and hardware. The security of the INMN and the IEDN is centrally defined with the Unified Resource Manager and enforced through HyperVisors and the TOR. All Virtual Servers and VLANs must pass through “access points” (HyperVisors and TORs) where their authorization is confirmed; the “access points” contain security enforcement that has been defined with Unified Resource Manager.

However, additional security controls are possible to define above the Hypervisor level, in the Operating System itself together with its applications! Understanding the level of security required and the isolation provided by the network virtualization management function of the Unified Resource Manager in collaboration with other firmware elements of the IBM zEnterprise System will help clients determine what, if any, additional security devices, like firewalls, are required in their enterprise solutions.

Built on this construct – zEnterprise – Innovation at every level

In the hardware management layer we have the System z server, and POWER based hardware. System x hardware and IBM appliances or optimizers (the specialty purpose processors versus the general purpose processors). Let’s ask ourselves … why should the hardware/firmware management be any different for each piece when they are all working together to provide the same business results? All have pieces in same business problem.

Moving into the Platform management layer – In the System z environment today customers manage the System z as an integrated and unified system – with single point of control for everything in the environment. But when you go into the distributed world you immediately get more risk. It’s just inherently more risky because every component has its own service and change methodology – there is one website for this piece and another website for another piece. As a customer, no matter how much you have control over your system – each piece of the solution is handled on an individual basis – with individual change control policies and nobody … except you (the customer) … has executed them in any given sequence. Every time something is done the customer takes the risk of doing something that isn’t compatible with the operating environment as a whole.

So this is what we want to help do with Unified Resource Manager. Take these many layers and extend them across the different architectures whether it’s the hardware (System z, Power, System x) or the platform (the virtualization and how it’s extended and managed).

IBM Systems Director and Unified Resource Manager both provide function at the hardware and platform management level. Use Unified Resource Manager for resources in zEnterprise.

Service Management – that’s Tivoli’s domain and this is fundamentally unchanged. And you need to understand that we are NOT talking about changes to the application environment. Anything working at the application level is unchanged.
There are four components of the network virtualization of the IEDN:

1. **Virtual Server (Virtual Server)** - This is where the application or service runs. It is a container that runs on a physical server.
2. **Firmware Configuration with Unified Resource Manager** - This is where the resources are defined for the virtual servers. It includes defining VLANs, VNICs, and VMACs.
3. **Network Virtualization Manager (NVM)** - This is where the network virtualization is configured and managed.
4. **Virtual Media Access Control (VMAC)** - This is the MAC address associated with the virtual network interface card (VNIC) used by the virtual server.

The Unified Resource Manager performs the virtual server definitions up to, but not including, the Operating System tasks. Therefore, Operating System tasks like installation of the software, applying fixes or patches, or backing up and restoring are performed as usual. For example, you might connect to the customer’s external network to perform these tasks.

After configuring the z196 as a node in an Ensemble, the “containers” for the Virtual Servers must be provisioned. Remember that the Virtual Server is wherever the Operating System is deployed. For example, an LPAR is a Virtual Server and z/OS might be deployed in it. Alternatively, Virtual Machines (or “Guests”) under z/VM could be Virtual Servers and Operating Systems might be deployed inside these Virtual Guests/Servers. The Virtual Server “container” provided by an Ensemble management represents the configuration required to define the Virtual Server – including LPARs, z/VM guests, Operating Systems, Blade guests – to the hypervisor. If the operating system definitions do not match the definitions of the Ensemble’s Virtual Server provisioning in the firmware, the Ensemble membership will not succeed. In z/VM the work at the HMC to define the Ensemble and its Virtual Servers provisions many of the definitions that are required to create Guests under VM, thus performing much of the work that a VM Systems Programmer would have to do in order to build a Virtual Guest. Some steps in the firmware configuration are explicitly executed by the administrator. Some occur transparently: MAC prefix assignment, MAC address assignment, connection of the Virtual Network Interface Card to the virtual network, etc. Some Operating Systems and Hypervisors – like z/VM and POWER – introduce more transparent provisioning of object definitions than others.

Notice the dotted line between one of the Blade Optimizers and the IEDN; notice the absence of a connection to the IEDN by the other Blade Optimizer. Depending on the type of blade deployed, you may or may not need to provide a connection between the optimizer and the IEDN.

There are four components of the network virtualization of the IEDN:

- **VLAN (Virtual LAN)** — A logical local area network that flows across the IEDN. A name and a numeric VLAN IDentifier define a VLAN. The TOR enforces membership in the VLAN.
- **VSwitch (Virtual switch)** — A virtual switch is a hypervisor component that provides virtualized network resources to a virtual server. Conceptually a VSwitch is represented for a z/OS LPAR by the OSX port. In z/VM the Hypervisor is a software VSwitch. AIX on a POWER blade also deploys a VSwitch (or VBridge) to interconnect Power LPARs to the hypervisor.
- **VNIC (Virtual network interface card)** — The Virtual Network Interface Card is the network resource that a virtual server uses to access the IEDN. The Virtual Network Interface Card is defined in the hypervisor through a VSwitch. (Again, conceptually the OSX port assumes the role of the VSwitch for z/OS.)
- **VMAC (Virtual media access control)** — The MAC address on a physical network card. A specific MAC address for an ensemble network resource is not assigned – only a specific MAC prefix is assigned. zManager assigns prefixes to the z/VM hypervisor and OSX. When servers initialize connections on either of these, a MAC is created and assigned to the server from either the hypervisor or OSX. The server can only send traffic using that MAC as its own. The enforcement points (i.e. OSX and z/VM) are each assigned a 3-byte MAC prefix from which a full 6-byte MAC address is created and assigned to QDIO connections. Access to the intra node management network is restricted to authorized management applications, and is only available through Port 0 of any OSA-Express3 CHPID configured with type OSM. Port 1 is not available for these communications.

VMACs on blades: The blade Hypervisors may obtain a MAC prefix either from a Unified Resource Manager definition or the blade Hypervisor may be implemented with a different MAC assignment scheme.

- Connectivity to the intra node management network is restricted to stacks that are enabled for IPv6.
- Connectivity to the intra node management network and to the intra ensemble data network is allowed only when zEnterprise is part of an ensemble.

**Dependencies:**

- This function is limited to OSA-Express3 ethernet features configured with CHPID types of OSX and OSM running on a zEnterprise. Also limited to OSA-Express4S configured with CHPID type of OSX. See the 2817DEVIC Preventive Service Planning (PSP) bucket for more information.
- This function is dependent upon the z/OS LPAR participating in a zEnterprise ensemble. See System zEnterprise Ensemble Planning Guide for more information.
z/OS at V1R12 with the appropriate APARs supports the Ensemble environment; APAR OA32099 (VTAM) and APAR PM0834 (TCP/IP) provide NEW FUNCTION for z/OS V1R10 and z/OS V1R11 to support TCP/IP in an Ensemble. See URL:
http://www.ibm.com/support/docview.wss?uid=isg1OA32099&myns=swgother&mymp=OCS55S3L&mync=E and
D385262FF0B0560905&d=OA32099&xref=%20. Consult PTFs UA57078 (Release 1A0) and UA57079
(Release 1B0). Minimum OSA levels are http://www-01.ibm.com/support/docview.wss?uid=swg27019754
(R10) and http://www-01.ibm.com/support/docview.wss?uid=swg27019755
(R11)

z/OS and VTAM: A VTAM Start Option enables ensemble participation by the TCP/IP stacks running in z/OS. The Start Option is "ENSEMBLE=YES\(NO\)." The ENSEMBLE setting is used to either permit or deny connectivity to the intraensemble data network and the intranode management network by allowing or denying activation of OSX and OSM interfaces. When the zEnterprise node is configured at the HMC as a member of an ensemble, you can change the value of the ENSEMBLE start option with the MODIFY VTAMOPTS command while VTAM is running.

Result: The ENSEMBLE setting is used to either permit or deny activation attempts for OSX and OSM interfaces. Modifying the ENSEMBLE start option from YES to NO does not cause z/OS Communications Server to take action on any active OSX or OSM interfaces.

The other operating systems all have their own definition types to configure what needs to match in the Ensemble Configuration that was performed at the HMC with Unified Resource Management.

Access to the intranode management network is restricted to authorized management applications, and is only available through Port 0 of any OSA-Express3 CHPID configured with type OSM. Port 1 is not available for these communications.

- Connectivity to the intranode management network is restricted to stacks that are enabled for IPv6.
- Connectivity to the intranode management network and to the intraensemble data network is allowed only when zEnterprise is part of an ensemble.
- This function is limited to OSA-Express3 ethernet features configured with CHPID types of OSX and OSM running on a zEnterprise. See the 2817DEVICE Preventive Service Planning (PSP) bucket for more information.
- This function is dependent upon the z/OS LPAR participating in a zEnterprise ensemble. See System zEnterprise Ensemble Planning Guide for more information.

Footnote 1: zTPF, zVSE, and Linux on z can all run native in an LPAR. However, in such a configuration they cannot participate in an Ensemble.
Only authorized applications and users can use the INMN. We show you how to do this for z/OS. Other operating systems have their own definitions and controls to authorize access to the INMN.

The SERVAUTH class EZB.OM.sysname.tcpname controls ability to access the intranode management network using OSM interfaces.

The NETACCESS controls are defined for z/OS with the SERVAUTH class EZB.NETACCESS.sysname.tcpname.security_zonename. This SERVAUTH class controls local user inbound and outbound access to network resources, and local user access to local IP address when explicitly binding to local interface (or using job-specific or destination-specific source IP addresses).
Networks in the Ensemble
Multiple Networks in a System z Configuration

- **Private system control network** (Internal)
- **Customer managed management network**
- **Customer managed external data network**
- **Intranode management network** (INMN within the node)
- **Intraensemble data network** (IEDN in the Ensemble)

This visual is one that you will often see when hearing about Ensemble Networking. It depicts 5 different networks that are maintained for a zEnterprise Node when it participates in Ensemble Networking, two of which are new to the concept of Ensemble Networking. We define some of these networks below, but we will separate out the networks and explain them in detail on the pages following this visual.

**Private system control network (PSCN).** A private network that is set up by the Customer Engineer. There is always some kind of internal network in any System z box. It ties the HMC to the Support Element. This network ties together the hardware components and also extends into the zBX's Advanced Management Module (AMM). But even there, on the zBX, it is used to tie together the hardware components.

**Intranode management network (INMN).** A private 1000BASE-T Ethernet network operating at 1 Gbps that is required for the Unified Resource Manager to manage the resources within a single zEnterprise node. The INMN connects the Support Element (SE) to the z196 and to any attached zBX.

**Intranode management network TOR switch (INMN TOR switch).** This 1 Gigabit Ethernet (GbE) switch is available exclusively on the zBX Model 002. It provides connectivity to the intranode management network (INMN) supporting communication between the Support Element (SE) and the Ensemble resources. There are two 1 GbE switches available in the zBX Model 002.

**Intraensemble data network (IEDN).** A private 10 Gigabit Ethernet network for application data communications within an ensemble. Data communications for workloads can flow over the IEDN within and between nodes of an ensemble. All of the physical and logical resources of the IEDN are configured, provisioned, and managed by the Unified Resource Manager.

**Intraensemble data network TOR switch (IEDN TOR switch).** This 10 Gigabit Ethernet (GbE) switch is available exclusively on the zBX Model 002. It provides connectivity to the intraensemble data network (IEDN) supporting application data within an ensemble. There are two 10 GbE switches available in the zBX Model 002. This is zBX Model 002 feature #0608.

**Customer management network.** This is a private network on which the HMC resides and over which the administrators communicate to reach the HMC.

**Customer external data network.** This is the external network that connects the network clients to the resources in the zEnterprise and the zBX.
When you migrate to a zEnterprise, regardless of the addition of a zBX Model 002, you still need to maintain your existing networks:

1) The management network through which you access your consoles. This network is usually implemented on a Private LAN with limited access. If the management network needs to be reached remotely, then security for remote access is introduced with Virtual Private Networks, and other security technologies like authentication, authorization, encryption, etc. VLANs can be introduced to segment and protect parts of the network as well.

   NOTE: In reality, there are two parts to the HMC Network: The Private System Control network (PSCN) and the Customer Managed Management Network. The PSCN is the portion of the network between the BPH and the Support Element (SE) and is internal to the System z. The Customer Managed Management Network is the portion of the network between the BPH and the HMC. For the sake of simplicity, we refer only to the Customer Managed Management Network.

   There is a maximum distance of 100 meters between the HMC and a switch to which the HMC is connected network

   The Customer-managed data network for the INTRANET and the INTERNET. These networks often connect through HiperSockets, through OSA ports (usually configured as CHPID Type OSD or OSE), or even through SNA-attached networks via a CHPID type of OSN or through channel-attached Front-end processors. VLANs can be introduced to segment and protect parts of the network as well.

2) And then, if you choose to implement Ensemble networking, you must also support two more networks: The intranode management network (INMN) and the intraensemble data network (IEDN). (See next page.)
For reliability in an Ensemble, you must order redundant switches and redundant OSA-E3 cards to attach to the switches in order to interconnect the members of an Ensemble. You must also provide definitions that secure the access to the two networks depicted: the intranode management network (INMN) and the intraensemble data network (IEDN). The switches reside in the FIRST RACK of a zBX.

**Intranode management network (INMN)**

- The Private System Control Network is the old term for the HMC/SE network. The INMN is an extension of this internal-only network.
- **For each z196 that participates in an Ensemble, define GbE ports of an OSA-Express3 1000Base-T card as CHPID Type of OSM in the IOCDS:** the OSM ports connect to the Intranode Management Network (INMN) over which the Unified Resource Manager defines, accesses, and manages the members of the ensemble. You can define ports that are shared among multiple logical partitions (LPARs) or ports that are dedicated to a single LPAR. A dedicated port is not required. It is recommended that you define ports that are shared just between the LPARs that work with your IBM BladeCenter Extension. **NOTE:** You do NOT define devices, links, or interfaces to this INMN from any of the Virtual Servers; this wholly self-contained private network dynamically builds the connections to the INMN when the server becomes a member of the Ensemble.
- **Note how the Support Element is still connected to the BPH switch as with the z10; however, now the OSM CHPID is also attached to the BPH Switch.**
- **HMC security is implemented with standard practices, but there are also additional safeguards for security, because the IP-V6 network is automatically created without a chance of human error during device definition. This is an isolated network that uses link-local addresses only; further authentication and authorization are implemented through the Firmware and through Operating System enablement to restrict access to the INMN.**

**Intraensemble data network (IEDN)**

- **10 Gigabit OSA-Express3 — QDIO (CHPID Type OSX):** Cables are 3.2 meters long from OSM to BPH in CEC and 26 meters from BPH to TOR.
- **Maximum of 16 data paths (8 pairs of redundant paths)**
- **Security is implemented with standard practices PLUS additional security mechanisms:** Layer 2 physical network, VLAN ID enforcement, access control, authentication, authorization, application security, routing table restrictions, IP Filtering, etc.
- **Networks can be further isolated using VLAN and VMAC segmentation of the network connections.**

---

**NOTE:**

- You must order redundant switches and redundant OSA-E3 cards to attach to the switches in order to interconnect the members of an Ensemble. You must also provide definitions that secure the access to the two networks depicted: the intranode management network (INMN) and the intraensemble data network (IEDN).
- The switches reside in the FIRST RACK of a zBX.
- **Intranode management network (INMN)**
  - The Private System Control Network is the old term for the HMC/SE network. The INMN is an extension of this internal-only network.
  - For each z196 that participates in an Ensemble, define GbE ports of an OSA-Express3 1000Base-T card as CHPID Type of OSM in the IOCDS: the OSM ports connect to the Intranode Management Network (INMN) over which the Unified Resource Manager defines, accesses, and manages the members of the ensemble. You can define ports that are shared among multiple logical partitions (LPARs) or ports that are dedicated to a single LPAR. A dedicated port is not required. It is recommended that you define ports that are shared just between the LPARs that work with your IBM BladeCenter Extension. **NOTE:** You do NOT define devices, links, or interfaces to this INMN from any of the Virtual Servers; this wholly self-contained private network dynamically builds the connections to the INMN when the server becomes a member of the Ensemble.
  - **Note how the Support Element is still connected to the BPH switch as with the z10; however, now the OSM CHPID is also attached to the BPH Switch.**
  - **HMC security is implemented with standard practices, but there are also additional safeguards for security, because the IP-V6 network is automatically created without a chance of human error during device definition. This is an isolated network that uses link-local addresses only; further authentication and authorization are implemented through the Firmware and through Operating System enablement to restrict access to the INMN.**

**Intraensemble data network (IEDN)**

- **10 Gigabit OSA-Express3 — QDIO (CHPID Type OSX):** Cables are maximum of 26 meters long from z196 to TOR & 10km long-reach (or up to 300m short-reach) to another zBX.
- **Security is implemented with standard practices PLUS additional security mechanisms:** Layer 2 physical network, VLAN ID enforcement, access control, authentication, authorization, application security, routing table restrictions, IP Filtering, etc.
- **Networks can be further isolated using VLAN and VMAC segmentation of the network connections.**
In the visual you see two zEnterprise nodes, each of which manages a zBX.

**HMC Communications:**
The HMC communicates with each node over the Customer Management Network by connecting through the Bulk Power Hub (BPH) switch which is attached to both the Support Element (SE) and the INMN of each node. The same HMC is used to manage all nodes within the Ensemble using the same network connectivity. The maximum distance that the HMC can be from the BPH is 100 meters because the connectivity is over 1000Base-T copper. In fact, all the HMCs and all the BPHs in all the z196s in the ensemble are on the same subnet.

**INMN Communications:**
The intranode management network interconnects only within a node; it does not span nodes. Therefore you see that there is an INMN for Node 1 and a second INMN for Node 2.

**IEDN Communications:**
The intraensemble data network spans nodes in the ensemble. Therefore you see a single IEDN interconnecting Node 1 and Node 2.

**Cabling Distances:**
The IEDN (zBX to zBX) distance can be up to 10km. The IEDN 10 km distance is fine for single mode fiber, but the maximum distance supported zBX to zBX on multi-mode fiber is 300 meters.

When there are multiple nodes in an ensemble, the IEDNs are interconnected via aggregation points (also called “Masters”) in the first and last zBX. This type of configuration is often referred to as a “Master-Slave” relationship. The first zBX maintains connections from one of the two 10 Gig switches to all the other members of the ensemble; the second of the two 10 Gig TOR switches in the last zBX (by convention called zBX #8) is the aggregation point back to all the ensemble nodes. A node without a zBX connects to another node’s 10 Gig TOR Switches, whose zBX is then aggregated as just described. (The 10-Gig TORs are connected to a management port in the 1-Gig TORs.)
In the visual you see two zEnterprise nodes, only one of which manages a zBX.

**HMC Communications:**
The HMC communicates with each node over the Customer Management Network by connecting through the Bulk Power Hub (BPH) switch which is attached to both the Support Element (SE) and the INMN of each node. The same HMC is used to manage all nodes within the Ensemble using the same network connectivity.

**INMN Communications:**
The intranode management network interconnects only within a node; it does not span nodes. Therefore you see that there is an INMN for Node 1 and a second INMN for Node 2.

**IEDN Communications:**
The intraensemble data network spans nodes in the ensemble. Therefore you see a single IEDN interconnecting Node 1 and Node 2. The OSXs of Node 2 are connected to ports within the IEDN portrange of J00-J07.
In the visual you see only ONE zEnterprise node. The second System z (either a z196 or another System z platform, which is not a member of the Ensemble) connects over a router (and firewall) from an OSD OSA to the TOR at a port within the portrange of J31-J39. Note that the Layer 3 Routing function would most typically be installed in a dedicated routing platform ("router") but could be in any node capable of terminating a connection in Layer 3 mode.
Intranode management network (INMN)

- For communication between the Support Element and the Hypervisors ONLY – not for Virtual Servers
- 1000Base-T OSA-Express3 (copper) — QDIO (CHPID Type OSM) — Cables from OSM to BPH are 3.2 meters long; from BPH to 1Gig TOR 26 meters long
- HMC security is implemented with standard practices PLUS additional security mechanisms:
  • Isolated IPv6 network with “link-local” addresses only; authentication and authorization and access control, etc.

For reliability in an Ensemble, you must order redundant switches and redundant OSA-E3 cards to attach to the switches in order to interconnect the members of an Ensemble. You must also provide definitions that secure the access to the two networks depicted: the intranode management network (INMN) and the intraensemble data network (IEDN).

The INMN is used for firmware management (platform management). It is not an application network at all and no application software uses this network. Once it is defined, it is essentially invisible to the users and the Virtual Servers; the Unified Resource Manager performs all the definition that is imbedded in the firmware; although there is an underlying VLAN, it is invisible and is not even defined through Unified Resource Manager panels. Virtual Servers communicate only to the SE over this network. One Virtual Server cannot even ping another Virtual Server over this network since each Virtual Server is isolated from the others. Besides the SE, only management applications (existing management applications and future ones when they become available) will be able to communicate with the Virtual Servers. (Management applications are called Guest Performance Agents and there are several others types of management applications as well.)

For each z196 that participates in an Ensemble, define GbE ports of an OSA-Express3 1000Base-T card as CHPID Type OSM in the IOCDS; the OSM ports connect to the intranode management network (INMN) over which the Unified Resource Manager defines, accesses, and manages the members of the ensemble. You can define ports that are shared among multiple logical partitions (LPARs) or ports that are dedicated to a single LPAR. A dedicated port is not required. It is recommended that you define ports that are shared just between the LPARs that work with your IBM BladeCenter Extension.

If the z/OS stack is enabled for IPv6, the stack defines two OSM interfaces. If the connectivity requirements on the previous chart are met, then Comm Server automatically starts these interfaces and dynamically creates TRLEs for them. These are IPv6 interfaces which only have a link-local address. These interfaces are always on a VLAN which is handled at the switch so the switch is unaware of the VLAN ID.

NOTE: You do NOT define devices, links, or interfaces to this INMN from any of the Virtual Servers; this wholly self-contained private network dynamically builds the connections to the INMN when the server becomes a member of the Ensemble. Generally speaking, z/OS as a member of an ensemble does not require a connection to the INMN. ZVM with Virtual Machines that are Virtual Servers MUST be connected to the INMN. (The z/VMS connection to the INMN is required even if any of the Virtual Guest are loaded with z/OS.) All members of the Ensemble MUST be connected to the intraensemble data network (IEDN).

Note how the Support Element is still connected to the BPH switch as with the z10; however, now the OSM CHPID is also attached to the BPH Switch.

HMC security is implemented with standard practices, but there are also additional safeguards for security, because the IPv6 network is automatically created without a chance of human error during device definition. This is an isolated network that uses link-local addresses only; further authentication and authorization are implemented through the Firmware and through Operating System enablement to restrict access to the INMN.

Cabling Specifications:
- zBX to zBX
- FC 0632 - Intraensemble Data Network 10Gb LR - Maximum distance between zBX to zBX with LR Optics is 10 km.
- FC 0633 - Intraensemble Data Network 10Gb SR - Maximum distance between zBX to zBX with SR Optics is:
  • 300 m for 50 micron at 2000 mHz-km
  • 82 m for 50 micron at 500 mHz-km
  • 33 m for 62.5 micron at 200 mHz-km
The INMN connections use the VLAN in **access mode** because only one VLAN ID is recognized on this internal management network.

The TOR switch handles VLAN tagging and the stack remains unaware of VLAN IDs for these interfaces.

The External Customer Data Network cannot connect directly to the intranode management network. The INMN is accessible only through the HMC. And the HMC is locally secured on a Private LAN with Authentication and Authorization. If being accessed remotely, the HMC is secured with Firewall Filtering, with a connection secured with Secure Sockets Layer (SSL), and further discrete authorizations. The Unified Resource Management functions are also secured with further discrete authorizations to its special functions.

The INMN uses hardware that is entirely different from the hardware being used by the external network or even the IEDN. There is physical separation. So there is no physical connection to the IEDN except to management ports which cannot forward traffic or receive traffic from data ports. (The 10-Gig TORs are connected to a management port in the 1-Gig TORs.)

An administrator can make use of the INMN if he uses ping, traceroute, diagnostics from an Operating System attached to the INMN, but such an administrator is allowed to issue such diagnostic commands only if authorized. Furthermore, such diagnostic commands can reach only as far as the Support Element, since each Virtual Server (including VSs in LPARs) is isolated unto itself over this INMN. Even the OSM OSA Port is operating in “Port Isolation Mode,” meaning that Virtual Servers in z196 LPARs cannot communicate with each other over the shared OSM ports.

**A Customer Requirement:** To access the INMN through a Network Management Application under secured conditions.

**Summary:** HMC security is implemented with standard practices, but there are also additional safeguards for security, because the IPv6 network is automatically created without a chance of human error during device definition. (You do not have to exploit IPv6 at all to create an Ensemble; the use of IPv6 is transparent to the user.) This is an isolated network that uses link-local addresses only; further authentication and authorization are implemented through the Firmware and through Operating System enablement to restrict access to the INMN. An IPv6 link-local addresses always start with FE:8, FE:9, FE:A or FE:B. Therefore a stack is able to recognize when an IPv6 address is link-local. A packet sent to a link-local address is sent to the local LAN via an interface that has a link-local IPv6 address assigned. This all occurs without the need to look up a route in a routing table. This type of routing is called **route-by-interface.** As a result, management interfaces in the INMN do not require any configuration of IP routes.

The INMN interfaces and their addresses are not reported to OMPROUTE nor can the operating system systems administrator add static or dynamic routes for these interfaces. Optionally you can specify a Security Class on the IPCONFIG6 statement of the z/OS TCP/IP profile (SecClass) to perform IPSec filtering. If desirable, you may allow stops, starts for these INMN interfaces or you may take packet traces and use OSAENTA on them.
The 1000Base-T OSA Configured as CHPID Type OSM

IOCP Example: You see here the IOCP statements to show what values the user should be entering on HCD/HCM panels. For example, in HCD you would be entering OSA-M for the TYPE and UNIT values. The illustrated syntax may not be valid IOCP input. For example, “Dynamic=Yes,Locany=Yes” would be entered on HCD/HCM panels, but are not valid IOCP parameters.

Notice first of all that this adapter is a four-port adapter and is also called a “multiport adapter.” This OSA-Express3 (1000Base-T) is engineered with two ports on each CHPID: CHPID x and CHPID y. The two ports on each CHPID are numbered port 0 and port 1. But note how the top half of the OSA-E3 is the mirror image of the bottom half with regard to the port number assignments; reading from top to bottom you see Port 0, Port 1, Port 1, Port 0. As with any OSA port, the portnames on the multi-port OSA-E3 must be unique to a CHPID in z/OS. Portname relief is implemented in z/VM and Linux on z, meaning that the portname is ignored by the OSA firmware whether specified in the Operating System definitions or not.

You may want to read more about this OSA type in technical document PRS3950 named “Migrating to a Multi-port CHPID OSA-E3: Avoiding Common Problems (CHPID Types OSD and OSE)” It is available at: www.ibm.com/support/techdocs.

You may also want to review Flash number FLASH10706 at the same web site; it is named “OSA-E3 Multiport and Portname Conflicts.”

Note (1): When an OSA CHPID is configured with a CHPID type of “OSM,” only PORT 0 is available; PORT 1 is not used at all in this case.

Note (2): The second CHPID on an OSA-E3 1000Base-T may be defined in the IOCDS as OSC, OSE, OSD, or even OSM. As per Note (1), if configured as OSM, PORT 1 is blocked from usage.

Operating System Definitions: QDIO TRLE definitions are always dynamically generated for connectivity to the intra node management network (CHPID type OSM) if the z196 has been enabled for Ensemble in the HMC, if the requisite Unified Resource Manager definitions have been created, if the stack has been enabled for IPv6, and, in z/OS, if the VTAM Start Option to enable Ensemble has been specified (ENSEMBLE=YES). For each dynamic TRLE for OSM, the TRLE name is IUTM0xx and the corresponding portname is IUTMP0xx (where xx is the value of the OSM CHPID). Up to 9 datapath devices are automatically generated.

Note: CHPARM=02 (default value) is the only valid value for an OSM CHPID. This disables priority queueing on the OSA port and supports a maximum of 1920 OSA valid subchannels (640 TCP/IP stacks). Specify either CHPARM=02 (default value) or CHPARM=00 for the OSD or OSX channel path. With CHPARM=00 an OSD or OSX channel path has priority specification enabled and supports a maximum of 480 OSA valid subchannels (160 TCP/IP stacks).

Security Controls: Besides the closed, Layer 2 (VLAN) implementation of this network, and the enforcement performed by the hypervisors for use of this network, some operating systems exercises additional security controls. In z/OS OSM usage is restricted to management applications with proper authorization. In order to establish a TCP connection over OSM, send or receive non-TCP traffic over OSM, or join a multicast group over OSM, an application must have READ access to the SAF resource named “EZB.OSM.sysname.tcpname.” On z/OS, OSM access control is mutually exclusive with network access control. In other words, if packets are sent or received over an OSM interface, network access control is not enforced, but OSM access control is enforced instead. Stack-generated ICMPv6 traffic (such as Neighbor Discovery) is exempt from the need for OSM access control. There are additional MLS considerations for OSM interfaces and authorized platform management applications using them. See the MLS chapter in the IP Configuration Guide for details on these.
Connection to the INMN network requires IPv6 support. If z/OS is not attached to the INMN network, it need not be enabled for IPv6, but this is not recommended. By not participating in the INMN network, z/OS loses access to the Platform Performance Monitoring capabilities of the management network.

Dual-mode stack (IPv4/IPv6) support is defined by using two NETWORK statements (one for AF_INET and one for AF_INET6) in the BPXPRMxx member.

This is considered an INET configuration, because only one stack is running in an MVS image, as is depicted on the previous page.

With the dual-mode stack, this visual shows that the IPv6 functions and protocols ICMPv6, NeD, MLD, and Autoconfig are automatically enabled.

ICMPv6 - The IP protocol concerns itself with moving data from one node to another. However, in order for IP to perform this task successfully, there are many other functions that need to be carried out: error reporting, route discovery, and diagnostics, among others. In IPv6, all these tasks are carried out by the Internet Control Message Protocol (ICMPv6). In addition, ICMPv6 provides a framework for Multicast Listener Discovery (MLD) and Neighbor Discovery (NeD), which carry out the tasks of conveying multicast group membership information (the equivalent of the IGMP protocol in IPv4) and address resolution (performed by ARP in IPv4).

Neighbor discovery is an ICMPv6 function that enables a node to identify other hosts and routers on its links. It corresponds to a combination of IPv4 protocols (ARP, ICMP Router Discovery, and ICMP Redirect). It maintains routes, MTU, retransmit times, reachability time, and prefix information based on information received from the routers. NeD uses Duplicate Address Detection (DAD) to verify the host’s home addresses are unique on the LAN. NeD uses Address Resolution to determine the link-layer addresses for neighbors on the LAN and Reachability Detection to determine neighbor reachability.

Multicast Listener Discovery (MLD) is the protocol used by an IPv6 router to discover the presence of multicast listeners (that is, nodes wishing to receive multicast packets) on its directly attached links, and to discover specifically which multicast addresses are of interest to those listeners. This information is then provided to whichever multicast routing protocol is being used by the router, in order to ensure that multicast packets are delivered to all links where there are interested receivers. MLD is derived from IGMPv2. One important difference to note is that MLD uses ICMPv6 message types, rather than IGMP message types.

IPv6 provides for both stateless and stateful autoconfiguration. Stateless autoconfiguration allows a node to be configured in the absence of any configuration server. Stateless autoconfiguration further makes it possible for a node to configure its own globally routable addresses in cooperation with a local IPv6 router, by combining the 48- or 64-bit MAC address of the adapter with network prefixes that are learned from the neighboring router. IPv6 allows the use of DHCPv6 for stateful autoconfiguration. DHCPv6 relies on a configuration server that maintains static tables to determine the addresses that are assigned to newly connected nodes. z/OS CS does not support DHCPv6.
Although INMN uses IPv6, the IPv6 usage in the INMN network is virtually transparent. IPv6 exploitation is unnecessary to create the Ensemble Environment.

You do not have to “learn” IPv6 to participate in an Ensemble. You only have to enable the stack to use IPv6, and there are only two changes you must make to do so: implement IPv6 using definitions in SYS1.PARMLIB(BPXPRMxx) and then optionally change any automated processes you may have to issue and interpret NETSTAT commands to utilize the LONG format of the command. The FORMAT LONG is used to support longer IPv6 addresses. Therefore, LONG FORMAT is always used when IPv6 is enabled. FORMAT SHORT is not supported when IPv6 is enabled. FORMAT can be defined in the IPCONFIG statement of a z/OS TCP/IP stack and thus cause all netstat commands to adopt the LONG format in an IPv4-enabled implementation.

Most Netstat Output in a TCP/IP Stack on z/OS that is not enabled for IPv6 can be displayed with either the LONG or the SHORT format. If you have automated operations that are triggered by NETSTAT Long Format messages under TSO (and TSO only), be aware of the fact that NETSTAT in LONG format does not produce Message Identifiers. Under TSO, Netstat output displays in IPv4 can contain messages with the Prefix “EZZ” as with EZZ2761I. These “EZZ” messages do not appear with other forms of the Netstat output, as under UNIX or with the MVS “D TCP/IP” variants of the Netstat command. Note that the Message Identifiers under TSO are displayed if the TSO user ID profiles are set to the value PROFILE MSGID and if the TCP/IP stack not enabled for IPv4 processing.

Note that “NETSTAT” is not regulated by standards; as a result each vendor platform can have a different implementation of the output displays for any of the netstat commands and options. Although you may have written shell scripts or Rexx programs to reformat output displays from a netstat command, every release of software or upgrade of an operating system can introduce changes that cause you to revisit your installation’s customized scripts to process netstat output. In other words, you are already familiar with this process of testing and possibly changing your customized scripts with every new release; the enablement of IPv6 is just another change that you must anticipate as usual.

One of the benefits of the netstat display output on z/OS is that, even with an IPv4 network, you can still choose to begin displaying netstat output using the LONG format. This means, that even before a migration to ensemble networking (which will require IPv6 enablement), you can begin the process of modifying your customized scripts and your automated operations that may have been relying on message identifiers.

If you have developed REXX programs that issue Netstat commands under TSO and parse the output lines based on message identifiers, you may need to change those REXX programs to use some other token in the output lines to decide the format of the line you are trying to parse. Netstat reports that display a new format when FORMAT LONG is used:

Netstat ALL/-A; Netstat ALLCONN/-a; Netstat BYTEINFO/-b; Netstat CONFIG/-f; Netstat CONN/-c; Netstat HOME/-h; Netstat PORTLIST/-o; Netstat SOCKETS/-s; Netstat (etc.).

Netstat GATE/-g will not be enhanced to support IPv6 routes. Use Netstat ROUTE/-r instead.

NOTE on z/VM: z/VM handles the INMN requirement for IPv6 differently. VM only supports IPv6 on a layer 2 mode Virtual Switch. The main TCP/IP stack does not have to talk IPv6 to be part of an ensemble. Z/VM has another internal z/VM Stack that is used for OSM connectivity. The customer does not use this stack. The netstat output for the IEDN is thus either IPv4 or IPv6, depending on whether the main stack is communicating using IPv4 or not. The customer will not have to configure this stack when Ensemble Managed.

NOTE: TELNET has a special set of display commands that also change their format once IPv6 is introduced. If you want to change the format before a migration to IPv6, just code FORMAT LONG in the TELNETGLOBALS statement.
Intraensemble data network (IEDN)

- 10 Gigabit OSA-Express3 — QDIO (CHPID Type OSX) — Cable = 26 meters — 10km**
- Security is implemented with standard practices PLUS additional security mechanisms: Layer 2 physical network (flat network), VLAN enforcement, access control, authentication, authorization, application security, routing table restrictions, IP Filtering, etc.
- Further isolation of Networks via VLAN and VMAC segmentation of physical network connections

** In a long-reach environment, the IEDN connection can be up to 10km long. However the actual limit on the distance is the distance between the HMC to the switch attached to the Support Element. (This distance is currently around 200meters.)

The intraensemble data network (IEDN) carries the data traffic among members of the ensemble. This network may be defined in the Operating System as an IPv4 or an IPv6 network. If you have not yet migrated your existing Customer Network to IPv6 you will probably for the short term continue to use IPv4 addressing in the IEDN. If you choose to implement IPv6 on the IEDN, you will want to understand the basics of networking with IPv6: IPv6 addressing, IPv6 protocol headers, IPv6 routing, IPv6 security, and so forth.

Whichever type of IP addressing protocol you choose to implement, you will discover that the intraensemble data network relies on Layer 2 VLAN routing. All members of an ensemble with a requirement to communicate with each other must belong to the same VLAN. If they are on the same VLAN, they then also all belong to the same IP Subnet. All network connections in an IEDN require that a Virtual Medium Access Control address (VMAC) be assigned to the connection together with a VLAN ID. The VLAN ID range is 10 - 1030. 10 is used for the Default virtual network, although this value can be modified through the UI. The VLAN ID must be one not already in use for another virtual network of the IEDN. Any VLAN ID(s) used on the IEDN must be authorized at the HMC. The TOR for the IEDN ports to the LPARs are defined in Trunk Mode.

All Operating System TCP/IP stacks require a Layer 3 routing table, but this table can be very simple, because the basic implementation of the IEDN uses a flat network, that is, one with all addresses in the same IP subnet. If you choose to implement some addresses in the members that are outside this IP subnet, or, if you elect to allow communication between the IEDN and your External Customer Network, you may need to implement more complex routing tables with static definitions. Although dynamic routing protocols may be deployed, not all network designs within the IEDN lend themselves to their use and you may want to avoid the use of dynamic routing protocols entirely within Ensemble. For example, OMPROUTE is not aware of the IEDN-to-IEDN forwarding restriction that z/OS enforces and could therefore calculate unusable routes. The IEDN-IEDN forwarding restriction could make its way into other operating systems in the Ensemble; therefore, the use of dynamic protocols should be discouraged.

For reliability in an Ensemble, you must order redundant switches and redundant OSA-E3 cards to attach to the switches in order to interconnect the members of an Ensemble. You must also provide definitions that secure the access to the intraensemble data network (IEDN).

SUMMARY:

Intraensemble data network (IEDN)

10 Gigabit OSA-Express3 --- QDIO (CHPID Type OSX)

Connected to authorized members of the Ensemble via the 10Gig TOR Switch; note that all physical switches are managed by the Support Element. (The 10-Gig TORs are connected to a management port in the 1-Gig management TORs.)

Maximum of 16 data paths (8 pairs of redundant paths)

There are eight OSA adapters (16 OSA ports) needed for maximum configuration in a node. Only the first pair of OSA cables is required to be connected to the managing zEnterprise.

Security is implemented with standard practices PLUS additional security mechanisms: access control, authentication, authorization, application security, routing table restrictions, IP Filtering, etc.

VLAN and VMAC segmentation of the network connections: Can assign Private network addresses (non-routable addresses) to the resources in the zBX, or can assign Public network addresses (routable) addresses to the resources, or can assign a mixture, especially if desiring to provide network reachability to a VIPA in the zBX.

Can implement with static routing. (Introduce dynamic routing only with a carefully designed dynamic routing environment.)
The 10 Gigabit OSA Configured as CHPID Type OSX  (z/OS)

IOCP Example: You see here the IOCP statements to show what values the user should be entering on HCD/HCM panels. For example, in HCD you would be entering OSA-X for the TYPE and UNIT values. The illustrated syntax may not be valid IOCP input. For example, "Dynamic=Yes,Locany=Yes" would be entered on HCD/HCM panels, but are not valid IOCP parameters.

Notice first of all that this adapter is a two-port adapter. Each port resides on a single CHPID. When one CHPID is configured as Type OSX, the other CHPID can be configured as either OSX or as OSD.

NOTE: You may want to read more about portnames that you can assign to these OSA ports in VTAM. Please see Flash number FLASH10706 (named "OSA-E3 Multiport and Portname Conflicts") at www.ibm.com/support/techdocs.

z/OS Operating System Definitions:

Dynamic Definition: When z/OS Communications Server participates in an ensemble environment, QDIO TRLE definitions are dynamically generated for connectivity to the intraensemble data network (CHPID type OSX) if the QDIO interface is defined with CHPIDTYPE OSX and the CHPID parameter. For each dynamic TRLE for OSX, the TRLE name is IUTXT0 xx and the corresponding portname is IUTXP0 xx (where xx is the configured CHPID parameter).

Manual Definition:

Tip: If the values that are chosen for the dynamically created TRLE definitions do not suit your needs (for example, you need more than 17 DATAPATH devices), you can define your own QDIO TRLE and configure the QDIO interface with the PORTNAME parameter. See INTERFACE -- IPAQENET OSA-Express QDIO interfaces statement and INTERFACE -- IPAQENET6 OSA-Express QDIO interfaces statement in z/OS Communications Server: IP Configuration Reference for more information.

Restrictions:

•If the QDIO interface that is used for connectivity to the intra ensemble data network is defined with the PORTNAME parameter, you must define a QDIO TRLE definition. The PORTNAME parameter value on the TRLE definition must match the PORTNAME parameter value. For information about restrictions on the TRLE name, see Transport resource list major node in z/OS Communications Server: SNA Resource Definition Reference.

•If z/OS Communications Server runs as a guest on z/VM and guest LAN definitions is used for connectivity to the intra ensemble data network, you must define the QDIO interface with the PORTNAME parameter.

Note: CHPARM=02 (default value) is the only valid value for an OSM CHPID. This disables priority queuing on the OSA port and supports a maximum of 1920 OSA valid subchannels (640 TCP/IP stacks). Specify either CHPARM=02 (default value) or CHPARM=00 for the OSD or OSX channel path. With CHPARM=00 an OSD or OSX channel path has priority specification enabled and supports a maximum of 480 OSA valid subchannels (160 TCP/IP stacks).

A note about MODEL on the IODEVICE. MODEL=model specifies the model number, if any, for the device. You can specify a maximum of four alphanumeric or special characters (#, @, or $) that represent the model number of the device. IOCP checks only the syntax for alphanumeric or special characters. IOCP does not validate the model value. For VM the MODEL=model parameter of the IODEVICE statement in the IOCPC input file has the same value as the MODEL=model parameter of the RDEVICE statement in the SYSTEM CONFIG file. For example, the following RDEVICE statement is in the SYSTEM CONFIG file: RDEVICE ... MODEL 11 The corresponding IODEVICE statement is: IODEVICE ...MODEL=11,...

IBM Reference Manuals:
Please consult the References at the back of this presentation to see a list of IBM Reference Manuals for System z and the zEnterprise hardware, for z/OS, and for z/VM. These references show you how to code for an Ensemble implementation.
The 10 Gigabit OSA Configured as CHPID Type OSX (z/VM)

OSA-E3 10 Gigabit

CHPID x

Port 0

osx

CHPID y

Port 0

IOCP

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=(09F0),* UNIT=OSA,UNITADD=00, * MODEL=X,DYNAMIC=YES,LOCANY=YES

Operating System Definitions (z/VM)

• Define controllers for the IEDN and INMN networks,
• Configure DirMaint™ (or other user directory manager), and
• Configure SMAPI servers.

• For more information, see “Configuring z/VM for an Ensemble” in z/VM: CP Planning and Configuration.

With z/VM the virtualization through the Unified Resource Manager panels performs many of the tasks that a VM System Programmer himself might otherwise do. DirMaint can be used as the Directory Manager. IBM provides the hooks for other directory managers to interface with the Systems Management API (SMAPI). Check with your software vendor to see if he has tested his directory manager in the zEnterprise ensemble environment.

IOCP Example: You see here the IOCP statements to show what values the user should be entering on HCD/HCM panels. For example, in HCD you would be entering OSA-X for the TYPE and UNIT values. The illustrated syntax may not be valid IOCP input. For example, “Dynamic=Yes,Locany=Yes” would be entered on HCD/HCM panels, but are not valid IOCP parameters.

Notice first of all that this adapter is a two-port adapter. Each port resides on a single CHPID.

NOTE: You may want to read more about portnames that you assign to these OSA ports in VTAM. Please see Flash number FLASH10706 (named “OSA-E3 Multiport and Portname Conflicts”) at www.ibm.com/support/techdocs.

IBM Reference Manuals:
Please consult the References at the back of this presentation to see a list of IBM Reference Manuals for System z and the zEnterprise hardware, for z/OS, and for z/VM. These references show you how to code for an Ensemble implementation.

A note about SHARED OSX ports: If you are sharing an OSX OSA port among LPARs and wish to prevent certain devices on the OSX CHPID from being shared, take advantage of the NOTPART keyword on the IODEVICE definition in the IOCDS. An LP cannot access a device if the LP is not specified in the device candidate list for the device, even if the LP can access a channel path assigned to the device. The PART or PARTITION keyword specifies the LPs that are in the device candidate list. The NOTPART keyword specifies the LPs that are not in the device candidate list. For example, if a CSS has three LPs (LP1, LP2, and LP3) and you specify NOTPART=(LP2), LP1 and LP3 can access the device, but LP2 cannot. This capability may provide assurances that certain Operating System and TCP/IP stack images cannot access the device, even if they can access a channel path assigned to the device. However, if this level of assurance does not satisfy an auditor, you may still exploit the z/VM or z/OS ISOLATE keyword to enhance the segregation of the Operating Systems sharing the OSX OSA. ISOLATE by itself may satisfy a security auditor or you may use ISOLATE in the interface definition together with the NOTPART definition in the IOCDS. However, only in very rare circumstances should you ever have to use ISOLATE in an ensemble environment, because the VLAN and VMAC enforcement of the ENSEMBLE already segregate traffic flows across the shared OSA. Implementing ISOLATE could, in fact, introduce network problems if not properly planned for.
TCP/IP Profile: IPCONFIG6 Statement

IPCONFIG6

- OSMSECLASS 255

>IPSECURITY

- OSMSECLASS security_class

- If desired to apply a Security Class to OSM, add above value to the IPCONFIG6 statement
- The value is meaningful ONLY if the stack has been configured with IPSECURITY

SAF Controls

- RDEFINE SERVAUTH EZB.OSM.sysname.tcpname UACC(NONE)
- PERMIT EZB.OSM.sysname.tcpname CLASS(SERVAUTH) ID(userid) ACCESS(READ)

OSM usage is restricted to management applications with proper authorization. In order to establish a TCP connection over OSM, send or receive non-TCP traffic over OSM, or join a multicast group over OSM, an application must have READ access to this SAF resource. OSM access control is mutually exclusive with network access control. In other words, if packets are sent or received over an OSM interface, network access control is not enforced, but OSM access control is enforced instead. There are additional MLS considerations for OSM interfaces and authorized platform management applications using them. See the MLS chapter in the IP Configuration Guide for details on these.

For OSM, VTAM dynamically creates TRLE with up to nine DATAPATH devices. You can see the dynamically created TRLE names and PORTNAMES on this chart. These TRLEs always use PORTNUM 0. VTAM looks for the smallest consecutive even/odd pair of devices to use for READ and WRITE. All other defaults are the same as for DynamicXCF HiperSockets:

- LNCTL = MPC
- MPCLEVEL = QDIO
- MPCUSAGE = SHARE
- MAXBFRU = 2
- MAXREADS = 2
- PACKING = OFF
- REPLYTO = 30
- STORAGE = DS
- LASTRW = DISALLOW

The IPCONFIG6 statement has a parameter which defines a security class for IP filtering for the automatically created OSM interfaces. This parameter is only meaningful if IPSECURITY is enabled for IPv6. This allows you to create a filter rule to permit traffic for OSM interfaces. This is similar to the SECCLASS parameter under DYNAMICXCF which covers dynamic XCF interfaces.

Interface names for the dynamically generated OSM Interfaces are: EZ6OSM01 or EZ6OSM02.

Several of the attributes for an OSM CHPID and Interface are different from those associated with the OSD Defaults: these attributes that are different are listed here:

- INBPREF DYNAMIC
- ISOLATE
- SECCLASS (value from IPCONFIG6 OSMSECLASS)
- VMAC (OSA-generated VMAC with ROUTEALL)

Security Controls: Besides the closed, Layer 2 (VLAN) implementation of this network, and the enforcement performed by the hypervisors for use of this network, some operating systems exercises additional security controls. In z/OS OSM usage is restricted to management applications with proper authorization. In order to establish a TCP connection over OSM, send or receive non-TCP traffic over OSM, or join a multicast group over OSM, an application must have READ access to the SAF resource named “EZB.OSM.sysname.tcpname.” On z/OS, OSM access control is mutually exclusive with network access control. In other words, if packets are sent or received over an OSM interface, network access control is not enforced, but OSM access control is enforced instead. Stack-generated ICMPv6 traffic (such as Neighbor Discovery) is exempt from the need for OSM access control. There are additional MLS considerations for OSM interfaces and authorized platform management applications using them. See the MLS chapter in the IP Configuration Guide for details on these.

IBM Reference Manuals:
Please consult the References at the back of this presentation to see a list of IBM Reference Manuals for System z and the zEnterprise hardware, for z/OS, and for z/VM. These references show you how to code for an Ensemble implementation.
Communications Server only allows OSX and OSM connectivity when both the z196 and the LPAR are configured as a member of an ensemble. An ensemble environment is only possible on a z196 machine. You can configure the z196 to be a member of an ensemble at the Hardware Management Console (HMC) using the Unified Resource Manager screens. To configure the LPAR as a member of an ensemble, specify the new VTAM start option ENSEMBLE=YES. Additionally, for OSM, the stack must be enabled for IPv6.

This shows the syntax for the VTAM start option. This option indicates whether or not this z/OS LPAR will participate in the ensemble. If the z196 is not configured as a member of an ensemble, then VTAM ignores this start option and issues existing message IST448I. The ENSEMBLE Start Option can also be altered with the MODIFY command. Please note however the following: “Modifying the ENSEMBLE start option does not cause z/OS Communications Server to take action on active OSX or OSM interfaces.”

Internal OSA interfaces require less configuration. Communications Server provides dynamic interface and TRLE definitions in many cases. Operational functions such as Start and Netstat are very similar to OSD. Note how both the VTAM TRLEs and the TCP/IP Interface definitions for the connections to the INMN’s OSM OSA ports are dynamically created for you. Then notice that even the VTAM TRLEs for the connections of the IEDN’s OSX OSA ports can be dynamically created. (You will see on the next page that you may also choose to predefine the TRLEs for the IEDN connections.)

For OSM CHPID Types, the stack defines two OSM interfaces as long as the z196 is enabled for the Ensemble, as long as the stack is enabled for IPv6, if VTAM’s Start Options include ENSEMBLE=YES, and if the OSX CHPIDs are attached to the partition. Communications Server dynamically creates TRLEs for these interfaces, dynamically creates IPAQENET6 INTERFACE definitions for the OSM connections, and automatically starts them. These are IPv6 interfaces which only have a link-local address. These interfaces are always on a VLAN; the TOR Switch handles the VLAN tagging, meaning that the connections are handled in ACCESS Mode at the TOR Switch. As a result, the stack is unaware of the VLAN ID. The stack does not report OSM interfaces to OMPROUTE and does not allow any routes to be added for them. Since the interfaces are not configured, but are subject to IPSec filtering, a security class may need to be configured for the OSM interfaces. Therefore, an optional SecClass configuration is allowed on the IPCONFIG6 statement. These OSM interfaces may be stopped and started, and are subject to packet trace and OSAENTA tracing as are any other OSA interfaces.

For OSX CHPID Types, VTAM dynamically creates a TRLE with up to 17 DATAPATH devices. The dynamically assigned names for the TRLE and for the OSX Portname are: TRLE name = IUTXT0xx (where xx = CHPID) and PORTNAME = IUTXP0xx. If you need more than 17 DATAPATH devices, you can predefine your TRLEs. All other defaults are the same as for DynamicXCF HiperSockets: LNCTL = MPC MPCLEVEL = QDIO MPCUSAGE = SHARE MAXBFRU = 2 MAXREADS = 2 PACKING = OFF REPLYTO = 30 STORAGE = DS LSTRW = DISALLOW

Alternatively you may choose to create your own TRLE definitions and assign your own portnames. You may wish to predefine the TRLE definitions if you are operating as a GUEST under z/VM.

IBM Reference Manuals:
Please consult the References at the back of this presentation to see a list of IBM Reference Manuals for System z and the zEnterprise hardware, for z/OS, and for z/VM. These references show you how to code for an Ensemble implementation.
You configure OSX interfaces using the INTERFACE statement for IPv4 or IPv6. You have two options. You can specify the OSX CHPID in which case VTAM dynamically creates a TRLE similar to what happens for HiperSockets. Alternatively, you can configure a TRLE and specify the PORTNAME as you do for OSD. This can be useful in a VM guest LAN environment. For OSX, you must configure a VLANID which matches one of the VLANs for which the OSA is authorized. As with OSD interfaces, OSX interfaces may be stopped and started, and are subject to packet trace and OSAENTA tracing as any other OSA interfaces.

For OSX configured with CHPID, VTAM dynamically creates a TRLE with up to 17 DATAPATH devices. The dynamically assigned names for the TRLE and for the OSX Portname are: TRLE name = IUTXT0xx (where xx = CHPID) and PORTNAME = IUTXP0xx. If you need more than 17 DATAPATH devices, you can predefine your TRLEs. All other defaults are the same as for DynamicXCF HiperSockets: LNCTL = MPC, MPCLEVEL = QDIO, MPCUSAGE = SHARE, MAXBFRU = 2, MAXREADS = 2, PACKING = OFF, REPLYTO = 30, STORAGE = DS, LASTRW = DISALLOW.

The visual also shows the syntax for an INTERFACE definition for OSX (CHPIDTYPE OSX). (CHPIDTYPE defaults to OSD.) The OSX syntax diagram is similar to that of OSD, but has some different defaults from OSD:

1) For OSX, you may specify either CHPID or PORTNAME depending on whether the underlying TRLE has been dynamically or statically defined.
2)  **VLANID** is required
3)  Defaults to INBPERF DYNAMIC. (We recommend adding WORKLOADQ to the “INBPERF DYNAMIC” so that you obtain three input queues for the OSA adapters.) In addition to enabling QDIO IWQ, the INBPERF DYNAMIC setting also dynamically tunes the read-side interrupt frequency for the OSA-Express feature. Note that the V1R12 IP Configuration Reference manual has an error in it indicating that WORKLOADQ is also the default when it is not.
4)  Requires an OSA-generated VMAC; you may not specify your own VMAC value.
5)  You must specify a subnet mask for IPADDR if you are defining an IPv4 interface.
6)  PRIROUTER/SECROUTER/NONROUTER are not applicable to an OSX Interface definition.
7)  All other parameters – including ISOLATE|NOISOLATE – default to OSD defaults

This chart also contains an example statement which defines an IPv4 OSX interface for CHPID F1 because the OSX TRLE for this CHPID was dynamically created by VTAM when the ENSEMBLE Start Option was enabled. (If the VTAM TRLE was dynamically created, use CHPID=<number> on the INTERFACE statement in TCP/IP. Alternatively you may instead use the PORTNAME assigned dynamically by VTAM.)

You also see an example for an INTERFACE statement when the TRLE was predefined; this is a case when you must specify PORTNAME in the INTERFACE definition. (If the VTAM TRLE was defined in a VBUILD TYPE=TRL, use PORTNAME=<TRLEportname> on the INTERFACE statement.)

IBM Reference Manuals:
Please consult the References at the back of this presentation to see a list of IBM Reference Manuals for System z and the zEnterprise hardware, for z/OS, and for z/VM. These references show you how to code for an Ensemble implementation.
z/VM Virtualization Introduces Simplified Management

NOTE: Power Virtualization with the Unified Resource Manager is similar to provisioning with z/VM, in that many of the definitions for the blade are automatically built during Ensemble membership definition.

With z/OS you saw that the firmware defines a "virtual entity" or "container" with rules that the installed operating system must adhere to in order to join the ensemble.

Here, on this visual, you see that the z/VM Hypervisor accepts commands through the System Management API that result in the construction and definition of the “container” that will hold the Virtual server once it is installed. Thus, the actions initiated through firmware results in definitions like the entries made in a VM directory – for example, the virtual NIC definitions for guest VSwitch attachment. Once the “container” definition for the Virtual Server is in place, a System Programmer can install the Virtual Server’s operating system.

(Note: The provisioning of the Virtual Server and the Virtual Network is also much simplified with the POWER blades, with many of the POWER tasks being completed during Ensemble definition.)

"Ensemble managed" from a z/VM LPAR perspective requires ALL of the following:

That the hardware ensemble indicator from the CEC be on
That SMAPI be up and communicating with the SE (HMC)

Through these communications we receive the required configuration elements from the z/Manager

NOTE: ZVMLXAPP is the Ensemble Management Guest (required for Unified Resource Manager). An internal port of 55555 is used by the SMAPI for the Unified Resource Manager to receive asynchronous notifications and pass them on via the event stream.

There should be no conditions, including error conditions, that will force the z/VM hypervisor to "opt out" and thus prevent any defined devices on the OSM and OSX CHPIDs from being varied online

For example, an inactive SMAPI server would prevent a z/VM from becoming ensemble-managed.

For example, a nested VM Hypervisor is not permitted to be ensemble-managed.

IBM Reference Manuals:
Please consult the References at the back of this presentation to see a list of IBM Reference Manuals for System z and the zEnterprise hardware, for z/OS, and for z/VM. These references show you how to code for an Ensemble implementation.

In addition, for instructions on building and working with a SMAPI server, consult: z/VM Version 6 Release 1 Systems Management Application Programming (SC24-6234-01)
You have seen this slide before. It is here to remind you that the HMC functions for working with the Unified Resource Manager perform many tasks that otherwise a system administrator or system programmer would have to perform. Some operating systems exploit the virtualization functions more than others.
Networking Topologies in the Ensemble
The initial implementation of the Ensemble on System z includes attachment to the “real devices” of the INMN and the IEDN networks by:

- z/OS (Base of V1R12; via PTF for V1R10 and V1R11)
  - An Operating System
- z/VM (Base of z/VM V6R1 with Small Programming Enhancement – SPE - and higher)
  - A Hypervisor (like an “LPAR” and capable of partitioning)
- z/VM TCP/IP stack (Base of z/VM V6R1 with SPE and higher)

Legend: dotted lines (………………) indicate “real device” attachment

Both z/OS and the z/VM Hypervisor, each in its own partition, own the real addresses assigned in the IOCDS to OSX and the OSM CHPIDs. (See the dotted lines on z/OS and on z/VM Hypervisor.) z/OS1 and the z/VM Hypervisor may vary online and offline these device or interface types. You should allow the Unified Resource Manager to be in charge of all Ensemble resource activity and not the operating system operator.

- Guests that do not have the required device driver support (back level) cannot directly connect to a real OSX/OSM. But they can still connect to the IEDN through a VSwitch if they have been authorized and configured to use OSDSIM support. In this case a guest can initialize an OSD Virtual Network Interface Card on the IEDN VSwitch and be able to communicate over the VSwitch's OSX uplink port. Linux and other guests which do not support OSX device drivers can talk on the IEDN network via OSDSIM support of the z/VM VSwitch. Their interfaces look like standard OSD interfaces in this case and you define an OSD Network Interface Card for them.

**NOTE 1:** This visual is only meant to illustrate the concept of Real versus Simulated Devices in the Ensemble. The IEDN path between the LPARs can be used for communication within the Ensemble. (At GA1 the HiperSockets is not part of the managed environment and provides no access to the IEDN. At GA1 the HiperSockets connection is considered part of the external network environment.)

**NOTE 2:** This visual does not depict the special management guest that interfaces between the management VSwitch (dashed lines) and the OSM CHPIDs.
The z/VM Hypervisor can be managed by the Ensemble with Unified Resource Manager firmware.

Guests under z/VM on a VSwitch can participate in the Ensemble:
- Linux on System z (via OSDSIM support)
- z/VSE (via OSDSIM support)
- TPF (via OSDSIM support)
- z/OS (pre-V1R10 requires OSDSIM support)
- z/VM TCP/IP (V6R1 with SPE or higher)

**Any IPL'able System with the required device driver support can connect to a real OSX device with either a dedicated connection **or** through a z/VM VSwitch.**

At GA1 only z/OS and z/VM TCP/IP have OSM/OSX Device Driver support; there is an SOD for device driver support in Linux.

This visual now focuses on the guests under z/VM in an LPAR.

All of the VM Guests depicted can participate in the Ensemble. Of the six z/VM guests depicted (Linux on z, z/VSE, TPF, z/OS, z/VM TCP/IP, and the nested z/VM Hypervisor) that are participating via the VSwitch in the Ensemble, only z/OS and z/VM TCP/IP enjoy at GA1 the device drivers to be able to vary online and offline the Guest's OSX interfaces to the Operating System. In addition, they have Guest Platform Management Provider agents (“GPMP”) to be able to communicate with the Management Guest (ZVMMAPLX) over the INMN; the Management Guest communicates with the Support Element.

- The management network can be virtualized in multiple ways: either through Real Device Support, or through VSwitch (with or without OSDSIM). The Virtual Servers with GPMP Agents and on the VSwitch can communicate over the INMN with the z/VM Management Guest. If you have device driver support to communicate directly with the IEDN, you can do so, although most implementers would choose EITHER the VSwitch connection to the IEDN **OR** the dedicated connection to the IEDN, but not both.
- The entire Management Guest and Networks are instantiated without customer intervention by the interface between Unified Resource Manager and the SMAPI server in z/VM. The OSM and the INMN VSwitches are in port isolation mode. (That is, Virtual Servers cannot communicate with each other over the INMN connection.)
- Consult one of the appendices for more detail on z/VM and its Management Guest.
Remember: The entire Management Guest and Networks are instantiated without customer intervention by the interface between Unified Resource Manager and the SMAPI server in z/VM. (That is, Virtual Servers cannot communicate with each other over the INMN connection.)

This visual now focuses on the guests under z/VM in an LPAR and how the instantiation of the OSM connections looks “under the covers.” Again, the customer does not need to build the VSwitches that are shown in the diagram. The Unified Resource Manager definitions provide for these VSwitches and connections. Connections to the INMN require a Layer2 VSwitch and even z/OS supports a Layer2 VSwitch, but ONLY for the INMN connections. The z/VM Management Guest (ZVMMAPLX) itself is attached via a Layer 2 VSwitch to the OSM OSA ports. Full Port ISOLATION is provided on the INMN VSwitch.

The z/VM Management Guest and any Virtual Server with the device drivers for OSM and with a GPMP agent (“GPMP”) -- are connected to the z/VM VSwitch, which interfaces directly with the OSM OSA ports. Other operating systems, like Linux on z, can connect to the GPMP over their own Virtual Switch. In this fashion the Guest Platform Agents or the Guest Performance Monitor (“GPMP” in the diagram above report statistics to Unified Resource Manager).

The performance agent is not aware of whether an operating system resides in an LPAR or a Guest Virtual Machine. The performance agent reports on the OS instance. The agent works in either the standalone environment (native LPAR) or the virtualized environment (on a VSwitch). In GA1 the performance agent primarily monitors and reports on CPU usage.


Security controls on z/OS
A guest platform management provider on z/OS is controlled through the started task HVEMCA, in SYS1.PROCLIB. To configure the execution and security environment for the guest platform management provider, you can work with the z/OS security administrator to modify the sample z/OS Security Server RACF® commands in SYS1.SAMPLIB member HVEENV. Through the modified JCL in the HVEENV member, you can:

- Create a group profile and user profile for the HVEMCA started task.
- Authorize the HVEMCA started task to use the INMN.
- If you are using security labels (SECLABELs) in your environment, assign the user of the HVEMCA started task to the security label of SYSMULTI.

Then you will authorize the UNIX Owner of that program (HVEMCA) to the EZB.OSM.sysname.tcpname resource.
The z/VM Hypervisor can participate in the Ensemble.

Guests under z/VM on a VSwitch can participate in the Ensemble:
- Linux on System z (via OSDSIM support)
- z/VSE** (via OSDSIM support)
- TPF** (via OSDSIM support)
- z/OS (pre-V1R10 requires OSDSIM support; ***z/OS requires a Layer3 VSwitch)
- z/VM TCP/IP (V6R1 with SPE or higher)

- In general, guests can be connected to the IEDN through a Layer 2 or Layer 3 connection (either VSwitch or Dedicated).
- ** Use either a VSwitch or a Dedicated connection, but not both.
- *** z/OS, z/VSE, and TPF require a Layer 3 connection.
- LEGEND: *** indicates Layer 3 only

This visual now focuses on the guests under z/VM in an LPAR.

The Ensemble does not support "Nested Hypervisors."

The z/VM Hypervisor itself (in the lower left-hand corner of the visual) participates in the Ensemble using real device attachments. All of the VM Guests depicted can participate in the Ensemble. Of the six z/VM guests depicted (Linux on z, z/VSE, TPF, z/OS, z/VM TCP/IP) that are participating via the VSwitch in the Ensemble, only z/OS and z/VM TCP/IP enjoy at GA1 the device drivers to be able to connect directly to the OSX and OSM ports. (You would normally choose one way or the other and not both.) (Note the dotted lines.) Not all Guest systems under z/VM can reside on a Layer 2 VSwitch; some, like z/OS, can be deployed only on a Layer 3 VSwitch.
• z/OS Supports “multipath” (a Layer 3 technology for concurrent use of OSA Interfaces)
• z/VM plans to introduce support on the IEDN for “link aggregation” on VSwitch for Guest (a Layer 2 technology for concurrent use of OSA Interfaces)
• Linux and other guests which do not support OSX can talk on the IEDN networks via OSDSIM support of the z/VM VSwitch. (SET VSWITCH GRANT userid OSDSIM)

The zEnterprise and its zBX are known for their availability features, one of which is based upon redundant hardware. Redundant Hardware can provide failover capabilities, and, in some cases, bandwidth by allowing nodes to exploit multiple networking paths concurrently to transmit data. In the example above you see that z/OS has a layer 3 feature called “Multipath” that allows the concurrent use of multiple connections to the two OSM ports and the two OSX ports depicted. Likewise z/VM has a layer 2 feature called “Link Aggregation” on a VSwitch that also allows concurrent use of OSD OSA ports. The Ensemble implementation plans to include this Link Aggregation support for the two OSM and two OSX OSA ports depicted. The Top of Rack switches in the zBX are also planning to support Link Aggregation. Of course the operating system loaded into a Virtual Server of the Ensemble must also support the feature.

A Linux on System z/VM guest can connect to ANY physical network in two basic ways:
- Method 1 - Directly, such as through a direct connection to a physical OSA port that is connected to the physical network.
- Method 2 - Indirectly, through a connection from a Virtual Network Interface Card on the Linux on System z guest to a z/VM VSwitch. This switch in turn connects to a physical OSA port that is connected to the physical network.

If the physical network is the new IEDN, this places certain requirements on the Linux on System z guest or z/VM. To connect any z196 image directly to the IEDN, you must connect through CHPID type OSX. For Method 1, the Linux on System z guest must have support for CHPID type OSX. Its QDIO device driver must be updated to understand OSX. However, it is possible that your version of Linux does not have this capability.

If your version of Linux does not support OSX, you must use Method 2 to connect Linux on System z guests to the IEDN. Normally, a VSWITCH must reflect the OSA architecture of its external uplink OSA port. A VSWITCH with an OSX OSA uplink port must present OSX characteristics to all downstream ports of its guests.

An example is their virtual Network Interface Cards. A VSwitch connected to CHPID type OSX only allows Linux on System z guests with OSX support to connect through it to that CHPID. z/VM allows authorized Linux on System z guests and other authorized users to define a legacy QDIO type NIC. This NIC is coupled to an IEDN VSwitch. This tactic allows a Linux on System z guest that has only the traditional OSA CHPID support to connect to the OSX (IEDN) network.

The administrator must use the HMC User Interface (UI) to specify permissions for a given Linux on System z guest. The important permission is that the guest has only Tivoli Provisioning Manager for OS Deployment support. Also, the target Linux on System z guest userid must be authorized for this "compatibility mode" through the SET VSWITCH GRANT userid OSDSIM command.

**Storage:** For P blades and Fiberchannel connections, each blade has 2 ports each on 2 switches = 4 ports total=32Gb@8Gb each. Each blade's VIOS will use the paths in round robin fashion such that all 4 can be used simultaneously.
An IEDN uses a flat network design in which interface connections to the IEDN reside in the same IP subnet and therefore the same broadcast domain. To reach an address in such a network, Layer 3 routing through an intermediate routing device is unnecessary. The IEDN simply uses Layer 2 VLAN routing.

This visual shows you how you might be using Virtual IP Addresses (VIPAs) as connection destinations in the two LPARs depicted. The Virtual IP Addresses here are in the same IP Subnet as the IP addresses of the physical (or virtual) connections. (Same Subnet VIPAs are the easiest way to implement Virtual IP Addresses.) The OSx port has been coded at z/OS with VMAC ROUTEALL, meaning that the OSx OSA can perform ARP processing for the VIPAs in the same IP subnet as the Interface itself. In fact, INBOUND routing into z/OS takes place using the VMAC definition and there is no OAT lookup for the VIPA address in the OAT Table. (Please see a later chart that shows how processing occurs for OSA OATs with either VMAC ROUTEALL or VMAC ROUTELCL.) Examine the Virtual Servers in the blades of the zBX. The TCP/IP stacks have associated an IP address in the flat network with the virtual network interface card; they have also assigned another IP address (like a VIPA or an IP-alias) as a target address for some application in the blade.

Important:
Not all TCP/IP stacks have the capability to respond to ARPs for an additional address like a VIPA or an IP-alias; this visual is merely an illustration of how you might be able to exploit something like a VIPA on a Virtual Server while avoiding the use of Layer 3 routing.

Note that the flat network design only works when the hosts that own VIPAs can respond to ARPs for those VIPAs. (With ARP Offload, the OSA ports respond to the ARP requests based upon the addresses registered in the OSA ports.) z/OS, z/VM, Linux on z (Novell or Red Hat) and whatever is sent over a VSwitch can cause a response to an ARP request; however other platforms on LPARs and blades may not. Therefore, in this scenario, the addresses in each LPAR are registered in the OSA, and the OSA port can respond to an ARP request through ARP offload. If these conditions are met, then flat Layer 2 routing suffices for the z196 part of the network. Furthermore, the addresses in a single IP subnet belong to the same Layer 2 VLAN and are assigned a VLAN ID. The VLAN Tags are part of the LAN Frame Header. All IP addresses in the same IP Subnet and VLAN can reach each other without the need for Layer 3 routing that would otherwise have forced a hop through a router. If nodes are unable to respond to ARP requests for any addresses, then static routing statements must supplement the simple routing table that all IP stacks use. This might, for example, be the case with Virtual Servers in the zBX. (Since we discourage the use of layer 3 routing in the Ensemble so as to preserve the simplicity of the routing design, you may want to eliminate dynamic routing protocols in the IEDN, or, if introduced, then plan them very carefully.) For example, OMPROUTE is not aware of the IEDN-to-IEDN forwarding restriction that z/OS enforces and could therefore calculate unusable routes. The IEDN-IEDN forwarding restriction could make its way into other operating systems in the Ensemble; therefore, the use of dynamic protocols should be discouraged.
IPv4 Ensemble Addresses in several Subnets:
- Physical IP 192.168.1.0/24; Virtual IPs 172.16.n.0/24, 10.1.1.0/24
- All IPv4 Ensemble Interface Addresses in VLAN 10
- Layer 2 Flat Network with Layer 2 VLAN Routing
- To reach the VIPAs: a simple static routing statement in Virtual Servers

In this scenario, the Virtual IP Addresses are in an IP Subnet that is different from the IP addresses of the physical connections. For this scenario in which the VIPA addresses are in a subnet that is different from that of the Ensemble Interfaces, you could consider Layer 3 Static Routing definitions. Scenarios with a single VLAN in the IEDN easily lend themselves to dynamic routing protocols, but, as soon as multiple VLANs are introduced, very careful planning is required with dynamic routing since we discourage the use of layer 3 routing within the IEDN anyway. (Note that z/OS has disabled the ability to route from one IEDN VLAN into a second IEDN VLAN unless Sysplex Distribution into the IEDN has been deployed. OMPROUTE is not aware of the IEDN-to-IEDN forwarding restriction that z/OS enforces and could therefore calculate unusable routes. The IEDN-IEDN forwarding restriction could make its way into other operating systems in the Ensemble; therefore, the use of dynamic protocols should be discouraged.)
External load balancers continue to operate as they do now even when Ensemble networking is introduced into the z environment.

Currently Sysplex Distributor (SD) distributes traffic within the z/OS image, and beginning with z/OS V1R11, SD can also distribute to its first non-z/OS target: the DataPower appliance. With DataPower as a blade in a zBX, Sysplex Distribution will also be able to reach into the blades of a zBX and perform traffic distribution to these appliances.
Is Layer 3 IP Forwarding Permitted? IPCONFIG DatagramFwd

1. Local traffic terminates or begins in a node.
   1. Each TCP/IP stack is able to process data destined for itself.
   2. Routed traffic is received from one network and then routed to another network.
      1. Depending on whether Layer 3 IP Forwarding is enabled or not, a TCP/IP stack may or may not be able to forward a received packet. (Visuals 1 and 2)
      2. IPCONFIG DatagramFwd | NoDatagramFwd
   3. In z/OS a Sysplex Distributor Node always forwards to a target destination. (Visual 3)

Although z/OS is used as an example here, most of the discussion about IP Forwarding applies to all TCP/IP stacks.

A feature of TCP/IP routing is to enable IP forwarding or to disable it. IP forwarding permits the transfer of data between networks. This IP forwarding capability is enabled or disabled in the z/OS TCP/IP stack with the IPCONFIG statement "DatagramFwd" or "NoDatagramFwd."

You may enable IP Forwarding between Subnets in a TCP/IP stack with IPCONFIG DATAGRAMFWD.

The stack may be a router or an endpoint along a connection establishment path.

You may disable IP Forwarding in a TCP/IP stack with IPCONFIG NODATAGRAMFWD.

The stack may be the endpoint only for connection establishment.

Regardless of the coding of NODATAGRAMFWD or DATAGRAMFWD, a z/OS Sysplex Distribution stack may forward received packets to a Sysplex Distribution target node over an XCF path or a VIPAROUTE path.
Secure Segregation within an IEDN: No Layer 3 IP Forwarding

Yes

TCPIP1 (z/OS1)

IPConfig
DatagramFwd

OSD1

VLAN10

OSX

Network1

IEDN

No

TCPIP2 (z/OS2)

IPConfig
NoDatagramFwd

OSD1

VLAN10

OSX

Network1

IEDN

TCPIP3 (z/OS3)

IPConfig
DatagramFwd or
NoDatagramFwd

OSX1

VLAN10

OSX2

VLAN11

IEDN

1. DatagramFwd and NoDatagramFwd operate as usual when routing is to occur between a non-IEDN interface and an IEDN interface. (Visuals 1 and 2)
   1. In addition, you may also deploy IP Filtering or Firewalls to permit or deny traffic.
   2. Within an IEDN, only Layer 2 forwarding of traffic is permitted.
   3. Forwarding occurs only between interfaces with identical VLAN IDs.
   4. If forwarding between different VLAN IDs is desired, then, as usual in the TCP/IP architecture, Layer 3 routing is required. (Packets must be sent to a router for routing to a different VLAN.)
   5. However, z/OS does not permit Layer 3 routing between OSX or IEDN interfaces at all.

5. Other operating systems (z196 or zBX) may allow Layer 3 routing between VLANs in the IEDN.

A feature of TCP/IP routing is to enable IP forwarding or to disable it. IP forwarding permits the transfer of data between networks. This IP forwarding capability is enabled or disabled in the z/OS TCP/IP stack with the IPConfig statement "DatagramFwd" or "NoDatagramFwd."

As you see in Visual #1, as long as routed traffic is permitted with IPConfig DatagramFwd and with any potential IP Filtering, then traffic entering an Ensemble Virtual Server by means of a non-IEDN path may be routed over the IEDN OSX OSA port into the IEDN. Visual #2 shows you that IP forwarding is generally disabled in TCPIP2 and so traffic cannot be routed between the external network and the IEDN.

However, as the third visual shows, traffic may not be routed between separate Ensemble VLAN IDs by the z/OS Ensemble Member. This is true regardless of the coding of DATAGRAMFWD or NODATAGRAMFWD. In the IEDN only Layer 2 forwarding is permitted and this can occur only if the VLAN IDs are the same. If it is necessary to route between two separate VLAN IDs in the IEDN, then the layer 3 routing table must be invoked to route outside the IEDN, through a router there, and then back into the IEDN over another VLAN. (See examples later in this presentation.)

As described IP Forwarding has been disabled in z/OS, but other operating systems (z/VM, for example) in the z196 or in the zBX may indeed still permit IP Forwarding to be able to route between separate VLANs within the IEDN in the same stack.

**EXCEPTION:** Regardless of the coding of NODATAGRAMFWD or DATAGRAMFWD, a Sysplex Distribution stack may forward received packets to a Sysplex Distribution target node over an XCF path or a VIPAROUTE path even within the IEDN.
Connecting the Customer Data Network to the intraensemble Data Network

2. Enter through a Router connection to the TOR – Switch connections not permitted!

1. Enter through an OSD Connection attached to an Ensemble Member.

Virtual Server TCP/IP1 (z/OS1)

Virtual Servers

Linux 51
zVM
Linux 55

MAC Filtering

VLAN & VMAC Enforcement

OSD

OSX

Router L3

External Customer Data Network

Virtual Server 10B

VMAC-B

VLAN & VMAC Enforcement

IP Filtering

MAC Protect

Top of Rack

More Background on Security Services and Mechanisms:
The VLAN ID enforcement for any Virtual Server attached to an OSX works as follows: if a z/OS Native LPAR is on the OSX, then the OSX performs the VLAN ID enforcement; if the Virtual Server is under z/VM and attached to a VSwitch, then the VSwitch performs the VLAN ID enforcement. The hypervisors on the Blade of the Virtual Servers of the zBX perform VLAN ID Enforcement in the zBX.

Remember that Security protection is much more than just inserting a firewall along a path. It encompasses all layers of the IP Stack: Application Security Mechanisms (Access Control Lists, Userid and Password checking, mapping mechanisms), Transport Security Mechanisms (SSL/TLS, AT-TLS), IP Layer Security Mechanisms (IPSec, IP Filtering, Intrusion Detection Services, Network Address Tables), Data Link Control Security Mechanisms (MAC Address Filtering, VLAN Segmentation or Segregation), and many more mechanisms too numerous to mention here.

With regard to MAC Filtering, the Unified Resource Manager can define MAC filtering for external MAC Addresses and must define permitted VLAN IDs on external connections. (These are the connections established on the TOR’s “egress” ports to and from the external network.) The MACs within the IEDN are managed by the Network Virtualization Manager. All MACs are allowed that originate from within the IEDN (they are managed by NVM); z/VM VSwitch MAC Protect function for Layer 2 is by default for IEDN type VSwitches. This MAC Protect function enforces that a VLAN ID sent during guest link initialization (SETVMAC) matches with what has been assigned by the z/VM hypervisor. In addition, all SOURCE MAC addresses on egress frames from the guest are verified to insure that only the assigned VMAC for the guest is being sent on outbound data transfers. This eliminates any attempt by the guest to spoof its source MAC address.

Note on use of VMACs: When an Operating System on z is using layer 2, it performs an ARP and builds Ethernet headers with VMAC. z/OS does not use layer 2 and so ARP is handled by the OSA and the OSA builds the Ethernet header. On a VSwitch under z/VM, you might have a Linux guest using Layer 2 in this case the VSwitch builds the Frame Header with the VMAC in it. But, if the guest system is using Layer 3 on a VSwitch, then a frame header is not necessary and the packet is forwarded over the VSwitch using only the IP address.

If you decide to permit communication between the External Customer Data Network and the Ensemble, you can keep this path secure.

First, determine whether you want to do this.

Second, understand that the Ensemble contains its own enforcement points within the TOR (for external connections on the egress ports) and also within the Hypervisors. All Virtual Servers and VLANs must pass through the enforcement points – “access points” (Hypervisors and TORs) – where their authorization is confirmed; the “access points” contain security enforcement that has been defined with Unified Resource Manager.

However, you may continue to implement other security services within the Ensemble by exploiting traditional security mechanisms: IP Filtering (a firewall function), encryption, access control lists, userid and password authentication, etc.

Third, determine how you will secure the external connections. Be aware of the fact that the TOR performs VLAN ID enforcement for connections to servers outside the zBX that are not attached to an OSA OSA port. If an ISAOPT appliance is on the zBX, then the TOR also performs the VLAN ID enforcement. (You configure the authorized VLAN IDs at the HMC as part of the Network Virtualization infrastructure.)

Note that for security purposes a Layer 2 connection from the external network into the TOR is not supported – the connection must be using Layer 3 protocols. Note that the Layer 3 Routing function would most typically be installed in a dedicated routing platform (“router”) but could be in any node capable of terminating a connection in Layer 3 mode.

Bridge Protocol Data Units – BPDUs – at Layer 2 cannot be successfully exchanged between an external Layer 2 switch and the IEDN TOR. Spanning Tree Protocol – STP – messages that might be received from external switches are filtered out at the TOR. This together with an external Firewall is to protect the security of the IEDN network by avoiding VLAN ID collisions. For example, if a customer were to attach an external switch to the TOR, and BPDUs and STP messages were not being filtered out, a customer’s external VLAN ID might be the same VLAN ID used within the IEDN and thus mistakenly cause the interconnection of external VLAN segments to the IEDN VLAN segments, thus impinging the security of the IEDN.

Firewalls

Firewall in front of LPAR that is attached to an OSD OSA (External firewalls are usually “appliance-based” firewalls that – unlike “host-based firewalls” – are stateful.)

IP Filtering in Virtual Servers residing on the zBX (may or may not be stateful)

IP Filtering in z/OS Policy Agent (a “host-based firewall” that is not stateful)

IP Filtering with Proventia Intrusion Prevention Services for Linux on z (a “host-based firewall” that is not stateful)

Other IP Filtering mechanisms (could be stateful or not)

Firewall in front of TOR in External network (External firewalls are usually “appliance-based” firewalls that – unlike “host-based firewall” – are stateful)

SERVHLL1 Devices: NETACCESS CONTROLS for IEDN

MAC ADDRESS FILTERING at the TOR

MultiLevel Security
When we talk about the IEDN, we point out that it is a FLAT network, using Layer 2 Virtual LANs to forward traffic from one end of this FLAT network to the other end. We also point out how this "one-hop" configuration reduces the complexity of the network design by eliminating network equipment (routers, cables, administration, and so on). Finally, we emphasize how the reduced complexity of the network design leads to the elimination of the typical security vulnerabilities of a multi-hop network and the reduction of network latencies. Certain VLANs in the IEDN can extend one hop to the external Layer 3 router. (See VLAN ID 100 in the visual.) But IEDN VLANs cannot extend into the external customer network by connecting through an external switch. (See VLAN ID 99 in the visual.)

In the visual you see that -- for security purposes -- a Layer 2 connection from the external network into the TOR is not supported – the connection must be using Layer 3 protocols. (Bridge Protocol Data Units – BPDUs – at Layer 2 cannot be successfully exchanged between an external Layer 2 switch and the IEDN TOR. Spanning Tree Protocol (STP) messages that might be received from external switches are filtered out at the TOR. This together with an external Firewall is to protect the security of the IEDN network by avoiding VLAN ID collisions. For example, if a customer were to attach an external switch to the TOR, and BPDUs and STP messages were not being filtered out, a customer’s external VLAN ID might be the same VLAN ID used within the IEDN and thus mistakenly cause the interconnection of external VLAN segments to the IEDN VLAN segments, thus impinging the security of the IEDN.)

**IMPORTANT Reminder:** The TOR is not just any switch .. the switch ensures that possibilities for LAN collisions and for misconfiguration do not impinge on the security of the network because certain standard switch functions like the exchange of Layer 2 messages have been disabled. -- Which is another reason why the Juniper switch itself cannot be swapped out for a different switch -- Unified Resource Manager integrates with the Juniper switch so as to eliminate these Layer 2 security exposures and to provide a simplified configuration interface that is independent of the platform- and vendor-unique Graphical User Interfaces with which an administrator would normally have to deal. As a result, it is not important that an administrator be familiar with the configuration syntax of a particular switch brand. Due to this simplified GUI and its integration into the zBX, the TORs require very little configuration -- many of the functions are fixed and relieve the Ensemble administrators of typical switch tasks. The only configuration necessary is for securing the attachment to the external network through access control lists to VLAN IDs and to Virtual MACs.
Virtual Server Provisioning with POWER®:
• Many tasks (i.e., VIOS) are automated and hidden from view
• Similar to z/VM

IEDN for the P Blade:
• Redundancy and Failover with “Network Interface Backup” (NIB)
• Primary / Backup Partition

VMAC-A VMAC-B VMAC-C VMAC-D

Virtual Switch
Phyp (VIOS)

Blade

NIC0 NIC1 ESM*

As you have seen before, the Ensemble implementation exploits platform-specific features for redundancy and high availability and performance improvements like IEEE Link Aggregation Groups, NIC teaming, NIB, z/VM vSwitch and VIPAs.
This is a simplified picture for the implementation of the platform-specific features of the Power Blade. **NOTE:** Power Virtualization with the Unified Resource Manager is similar to provisioning with z/VM, in that many of the definitions for the blade are automatically built during Ensemble membership definition.
You see the zBX, one Power Blade, and you also see 4 partitions with Virtual Servers (VS) in them. The Virtual Bridge is like a VSwitch under VM, in that it represents the Hypervisor component that performs VLAN and VMAC enforcement. You do not see a 5th partition called the “VIOS Partition,” which is part of the Power Blade implementation; however, in the Ensemble, this partition is automatically generated with the Unified Resource Manager firmware definitions. In fact, just as with z/VM, the Unified Resource Manager definitions perform much of the work of provisioning the VIOS partition and the other components, leaving less to the system administrator. The two physical ethernet connections (NIC connections) to the Top of Rack Switches are associated with each other to provide Primary and Backup failover – not to provide multiple paths. The NICs are associated with Electronic Switch Modules (ESMs).

The **Unified Resource Manager completely controls the VIO server.** The Virtual Server does not have physical connectivity to the physical ethernet adapters; it has virtual connectivity through the VIOS to the physical adapters. The profile for the VIOS is predefined and you do not play a role in defining its characteristics; this is all performed out of sight by the entries in the Unified Resource Manager.

If you need more information on the Power Blade implementation details, please consult the following documentation:


NIC = Network Interface Card (the physical LAN Port)
VMAC = Virtual Network Interface Card (the Virtual Ethernet Port on each Virtual Server)
The NICs are connected to an Electronic (Ethernet) Switch Module (ESM) which is then connected to a TOR. “ESM” is a generic name; specific names are Blade Switch Module (BSM) for the 1Gig switches and High Speed Switch Module (HSSM) for the 10Gig switches.
NIC1-----ESM1-----TOR1 and NIC2-----ESM2-----TOR2
The traffic may be distributed equally through the NICs. For example, 50% of the Blades use NIC1 as backup and the other 50% use NIC2.
The Primary and Backup roles for exploiting two LAN connections have different names depending on the platform in which the feature is implemented:

- P supports a form of “Network Interface Backup” in order to provide redundant LAN connections for primary and backup only. On the blade in the Ensemble the actual implementation is an IEEE Link Aggregation group with one NIC and then one as a backup. The “pHyp” Hypervisor is PowerVM.
- X Blade support, for which there is an Ensemble Statement of Direction, supports “NIC teaming” or “NIC bonding” in order to provide redundant LAN connections for primary and backup only; the “xHyp” hypervisor (i.e., KVM) support for this feature exploits the bonding device driver.
This screen is concerned with the Virtual Switch that will be used for the local connections between the Virtual Servers and the VIOS of the P-blades only. P does not use a Link-local address on this VSswitch; it uses a “local unicast prefix” which is defined on the Unified Resource Manager screen devoted to “Network Details” about the ITSO Ensemble Details. Some parts of vios needed for communication with virtual servers can't handle IPv6 link local addresses. Therefore we use Unique Local Unicast addresses instead of IPv6 Link-Local interface addresses on this VSswitch. This requires the generation of a random IPv6 prefix. While it works the same as IPv6 Link Local, the hypervisor requires a display in Unified Resource Manager to handle rare collisions. Although the probability of a collision is very low (like the MAC prefix), we need a way to handle such situations. To avoid collisions (similar to the MAC prefix case) we provide a way for the customer to generate a different one on the Unified Resource Manager screen.
Implementation Planning for an Ensemble
Sample Implementation Plan: Workload Has Been Identified

1. Order the necessary manuals for the new environment
2. Identify the Workload eligible for an Ensemble Environment
3. Establish Naming Conventions for the Ensemble and its Members (Nodes)
4. Design and Document the Virtual Server environment of the workload
   - Designate the System z and the Blade platforms that are to run the workloads
   - Establish naming conventions for the Virtual Servers
5. Design and Document the Virtual Network connectivity for the Virtual Servers
   - Internal VLAN connectivity and naming/numbering conventions
     - for VLANs and IP Addresses
   - External Network connectivity
     - Through the zBX Tor
     - Through an Ensemble LPAR on the z196
6. Design and Document the Security environment of the Ensemble
   - User IDs, Resource Roles, Task Roles
   - External Firewall connectivity
   - Security implementations within the operating system
7. Size the design
8. Schedule a Design Review with your Team and external Subject Matter Experts
9. Order the necessary hardware and software
10. Document the Task Plan
    - Site Planning - Operating System Changes - Ensemble Tasks - Other

Many installations will know exactly which applications they would like to exploit with the zEnterprise System solution. For such installations, it is wise to use these application workloads as a basis for sizing and ordering the system components. Other installations may not know exactly which workloads they would like to implement with the zEnterprise System, but nevertheless know that their company growth and profile will provide an eligible workload in the short term. Such installations might order a minimal zBX configuration with licenses for the blade types that they are most likely to exploit.

This visual shows a sample implementation plan for a preordained workload.
The types of networks you are probably familiar with could look something like this conventional design.
The typical design you saw on the previous page could end up looking like this if it were to be implemented in the Ensemble.

When we talk about the IEDN, we point out that it is a FLAT network, using Layer 2 Virtual LANs to forward traffic from one end of this FLAT network to the other end. We also point out how this "one-hop" configuration reduces the complexity of the network design by eliminating network equipment (routers, cables, administration, and so on). Finally, we emphasize how the reduced complexity of the network design leads to the elimination of the typical security vulnerabilities of a multi-hop network and the reduction of network latencies. Certain VLANs in the IEDN can extend one hop to the external Layer 3 router. But IEDN VLANs cannot extend into the external customer network by connecting through an external switch.

The differentiators between the previous visual of a traditional network and this visual with the Ensemble Network depicted in this visual are:

1) The simplified, secured, and centralized configuration of the virtualized environment with the Graphical User Interface of the Unified Resource Manager;
2) The enforcement of the configuration and its security through the Hypervisors of the Virtual Servers;
3) The simplified configuration and security enforcement points of the Top-of-Rack switches that extend into the external network.

IMPORTANT Reminder: The TOR is not just any switch. The switch ensures that possibilities for LAN collisions and for misconfiguration do not impinge on the security of the network because certain standard switch functions like the exchange of Layer 2 messages have been disabled. -- Which is another reason why the Juniper switch itself cannot be swapped out for a different switch -- Unified Resource Manger integrates with the Juniper switch so as to eliminate these Layer 2 security exposures and to provide a simplified configuration interface that is independent of the platform- and vendor-unique Graphical User Interfaces with which an administrator would normally have to deal. As a result, it is not important that an administrator be familiar with the configuration syntax of a particular switch brand. As a result of this simplified GUI and its integration into the zBX, the TORs require very little configuration -- many of the functions are fixed and relieve the Ensemble administrators of typical switch tasks. The only configuration necessary is for securing the attachment to the external network through access control lists to VLAN IDs and to Virtual MACs.
This visual shows you that Virtual Router Redundancy Protocol (VRRP – RFC 3768) and Cisco Hot Standby Router Protocol (HSRP – RFC 2281) are valid in traditional network configurations and continue to be so when implemented in Ensemble networks. Both protocols provide router redundancy by configuring heart-beat messages between routers so that one router can take over a failed address from another router. Cisco’s HSRP backs up virtual IP addresses so that each router can take over the duties of the other router; VRRP backs up the real interface addresses. It is common to provide this type of backup on the “Access Layer” or periphery layer of the network in which the client platforms reside. In a conventional switching network, such protocols can less typically also be exploited at the “Core Layer” of a network topology, that is, between a router and a Server (or Mainframe) platform. With the Top-of_Rack (TOR) Switches of the zBX the same types of redundant configurations would continue to work.
What May Catch You by Surprise: Non-Disruptive Test Bed for IPv6

• Securely Isolate an IPv6 Test (or even Production) Network from the IPv4 network.

The visual shows you that you could isolate certain Virtual Servers in the IEDN and then use them for testing IPv6 network connectivity. In fact, this IPv6 test bed might be a practical workload for a company that has ordered a zEnterprise System but that has not yet identified a production workload to run in the Ensemble.
Network Virtualization Management: Secure Networking with VLANs in the Ensemble
VLANs are a means of building distinct security zones across the same physical network segment. Over the years there have been concerns in the industry about VLANs in terms of security, because both the MAC address and the VLAN ID can be spoofed or altered. However, in an IEDN these types of issues do not apply, because we use the Network Virtualization Management (NVM) features to control all access to the LAN via our hypervisors – including the OSX implementation as a Hypervisor VSwitch. The VLANs and the VMACs are controlled through NVM assignments. (With OSX an Operating System cannot build its own Logical Link Control – “LLC” – header.) The only exception to the security with a VLAN implementation is when an outside network connects to the IEDN via the external ports on a TOR; for this exception, the customer is responsible for securing the access.

Ensemble members that reside in a zBX may not be sharing ports at all. Even in such a situation, the VLAN ID prevents sending a message intended for one member to another. In this visual you see that Virtual Servers z/OS1 with TCPIP1 and z/OS3 with TCPIP3 cannot communicate with each other over the OSA port because they are attached to different VLANs. However, Virtual Server TCP1 can communicate via the OSX OSA port to the Virtual Servers in the zBX that are also attached to the same VLAN ID of 10 (Virtual Servers 10B and 10A). Likewise, Virtual Server z/OS 3 with TCPIP3 on VLAN 11 can communicate with the Virtual Servers in the zBX that are also attached to VLAN ID of 11 (Virtual Servers 11C and 11D).

Virtual Server Operating Systems should not forward from one VLAN ID to another on an OSX CHPID via Layer 3 routing. In fact, on z/OS IP Forwarding has been disabled when traffic comes in on an OSX port and wishes to exit on another (or the same) OSX port. Both z/OS and z/VM can implement the ISOLATE function for an OSD or OSX interface; although ISOLATE is not necessary to isolate traffic over a Shared OSA implementation when one has deployed different VLAN IDs (as in the diagram above), you can still code ISOLATE on the Interface definition to prevent communication over a shared OSA port.

When anything is connected to the IEDN, a VLAN must be used. When the traffic hits the TOR port, the sending adapter (server) must have applied a VLAN ID tag. Therefore, for z/OS it will be tagged by the OSA at z/OS, and for an external router that wishes to connect to the TOR, the traffic must be tagged by that router. VLAN ID enforcement adds a layer of security; it takes place at the Hypervisor. (hypervisor. A program that allows multiple instances of operating systems or virtual servers to run simultaneously on the same hardware device. A hypervisor can run directly on the hardware, can run within an operating system, or can be embedded in platform firmware. Examples of hypervisors include PR/SM, z/VM, and PowerVM.™) In fact, even the OSX is considered a Hypervisor VSwitch, or a type of PR/SM Hypervisor. z/VM is a also a Hypervisor for a direct attachment to the OSX, where z/VM itself is considered a type of PR/SM Hypervisor.)

In the scenario depicted, the TOR IEDN Ports to the LPARs would be configured in TRUNK Mode because the Operating Systems are defining their VLAN IDs (are “VLAN-aware”). If z/VM is a host for a Virtual Server whose Operating System is not VLAN-aware, the z/VM VSwitch – which is a Hypervisor – handles the VLAN ID on behalf of the Virtual Server. TOR ports to the Virtual Servers on the BLADEs are also configured in TRUNK mode. The TOR ports to any ISAOPT blades are configured in ACCESS Mode. The hypervisor – whichever it may be (OSX for native z/OS or z/VM, VSwitch for Virtual Servers on a VM VSwitch, or the Hypervisor on the blade) – performs the enforcement. The TOR performs the enforcement for anything that does not fall into these categories just mentioned. For example, if a native LPAR is communicating with a virtual server on a blade, the OSX will perform VLAN enforcement; the data passes through the TOR and the Electronic (Ethernet) Switch Module (ESM) without checking, but it is then checked again at the blade's VSwitch (=hypervisor).
Secure Segregation of VLAN 10 from VLAN11 (IEDN): 1 VSwitch

- If there is a z/VM VSwitch as depicted, the VSwitch, a Hypervisor component, prevents communication between the VLANs.
  - Layer 2 protocols cannot route between different VLANs; a layer 3 router is required to route between different VLANs.

There is VLAN ID enforcement on the z/VM VSwitch
Secure Segregation of VLAN 10 from VLAN11 (IEDN): 2 VSwitches

- If there are two z/VM VSwitches as depicted, the OSX VSwitch prevents communication between the VLANs.
  - Layer 2 protocols cannot route between different VLANs; a layer 3 router is required to route between different VLANs.

There is still VLAN ID enforcement on each individual z/VM VSwitch. But now you see that the OSX VSwitch will also not allow forwarding between two VLANs when the VLANs are on separate VSwitches.
Routing between VLAN IDs in an IEDN?  Maybe or Maybe Not!

Extra Security: You may not route between VLAN IDs within the IEDN unless you have a Layer 3 Router!

Only nodes that reside in the same VLAN can communicate with each other at Layer 2.
On z/OS you may not route between VLAN IDs within the IEDN at all: IP Forwarding from OSX to OSX is restricted.

You may have implemented multiple VLANs on the OSX OSA port; see VLAN10 and VLAN11 in TCPIP3 of the visual. This could allow members of the IEDN that reside in different VLANs to access the services of a DB2 Server; examine the arrows at the ends of the dashed lines in the visual. But you may ask yourself whether you are allowed to route between VLANs. The answer to this question depends on whether the IEDN’s Virtual Server operating system permits layer 3 routing between VLANs and on whether you have permitted routing through routers external to the IEDN.

Normally Layer 3 IP Routing can be used to route between nodes on different VLANs. (Layer 2 routing between different VLANs is not possible and is not part of the 802.1q standard.) However, in the ensemble for z/OS, the ability to invoke Layer 3 routing over an OSX CHPID has been disabled to further strengthen the segmentation of traffic in the Ensemble. This design is important for the following reasons.

Over the years it has become widely accepted that VLANs are not viable forms of security, because both the MAC address and the VLAN ID can be spoofed or altered. However, in an IEDN these types of issues do not apply, because we control all access to the LAN via our hypervisors – including the OSX implementation as a VSwitch. (With OSX an Operating System cannot build its own Logical Link Control – “LLC” – header which includes the VLAN ID.)

Ensemble members that reside in a zBX may not be sharing ports at all. Even in such a situation, the VLAN ID prevents sending a message intended for one member to another. In this visual you see that the IEDN secure architecture does not permit routing through the z/OS stack in order to interconnect nodes on VLAN10 and VLAN11. Likewise, in the zBX, Virtual Server 10B on VLAN10 cannot communicate with Virtual Server11D on VLAN11.

Recall that IP Forwarding has been disabled on z/OS when traffic comes in on an OSX port and wishes to exit on another OSX port. (The only exception is for Sysplex Distribution routing.) The z/OS architects strongly believe that routing between VLANs should be performed by an external router and ONLY WHEN FIREWALL IP FILTERING is used to permit such routing. (See later slides about routing between VLANs.) Other operating systems have not disabled the IP Forwarding function over IEDNs; however, even with these operating systems, such forwarding should be permitted only if FIREWALL IP FILTERING allows it. (Again, see later slides for examples of this.)
What if I Must Route between VLANs of the IEDN?  Example 1

**Extra Security:** You may route between VLANs through a non-IEDN router if necessary, but use a Firewall to maintain secure environment.

**Example:** non-IEDN router (Layer 2 not permitted)

---

When anything is connected to the IEDN, a VLAN must be used. When the traffic hits the TOR port, the sending adapter (server) must have applied a VLAN ID tag. Therefore, for z/OS it will be tagged by the OSA at z/OS, and for an external router that wishes to connect to the TOR, the traffic must be tagged by that router.

Virtual Server Operating Systems should not forward from one VLAN ID to another on an OSX CHPID via Layer 3 routing. In fact, on z/OS IP Forwarding has been disabled when traffic comes in on an OSX port and wishes to exit on another OSX port.

If you find it necessary to route between VLAN IDs by exiting to the external network to allow a router to route via Layer 3 back into the IEDN and another VLAN, then you will want to implement a Firewall to keep the traffic secure. Many security mandates require Stateful Firewalls, and most external Firewalls with IP Filtering do implement stateful packet inspection. Note that for security purposes a Layer 2 connection from the external network into the TOR is not supported – the connection must be using Layer 3 protocols. (Bridge Protocol Data Units – BPDUs – at Layer 2 cannot be successfully exchanged between an external Layer 2 switch and the IEDN TOR.)

You must also implement static routes to force Virtual Server11D to route to the external router depicted; the external Layer 3 router then routes back into the zBX to create the connection to Virtual Server10B. You must use static routing to accomplish this, because dynamic routing will have built the route between the two servers unless the Virtual Server implementation has blocked this internal routing over the IEDN as z/OS has done. The router must have tagged packets with the VLAN ID that is to be used within the IEDN. (That is, in our scenario here, the router will have tagged the traffic to the IEDN’s VLAN10 with VLAN10 and the traffic to the IEDN’s VLAN11 with VLAN11.)
Virtual Server Operating Systems should not forward from one VLAN ID to another on an OSX CHPID via Layer 3 routing. In fact, on z/OS IP Forwarding has been disabled when traffic comes in on an OSX port and wishes to exit on another OSX port.

If you find it necessary to route between VLAN IDs by exiting to the external network to allow a router to route via Layer 3 back into the IEDN and another VLAN, then you will want to implement a Firewall to keep the traffic secure. Many security mandates require Stateful Firewalls, and most external Firewalls with IP Filtering do implement stateful packet inspection.

You must also implement static routes to force Virtual Server11D to route to the external router depicted; the external Layer 3 router then routes back into the zBX to create the connection to Virtual Server10B. You must use static routing to accomplish this, because dynamic routing will have built the route between the two servers unless the Virtual Server implementation has blocked this internal routing over the IEDN as z/OS has done. The router must have tagged packets with the VLAN ID that is to be used within the IEDN. (That is, in our scenario here, the router will have tagged the traffic to the IEDN’s VLAN10 with VLAN10 and the traffic to the IEDN’s VLAN11 with VLAN11.)

The example above shows the static route coding at Virtual Server11D so that it can route through the external router that can be found at 10.11.11.200.

Equivalent coding at Virtual Server10B would be:
ROUTE 10.11.11.11/24 10.10.10.200 eth0 mtu 8992

The router would have the necessary coding to be able to route back into the zBX to reach the desired destination.
What if I Must Route between VLAN IDs of the IEDN? Example 2

**Extra Security:** You may route between VLAN IDs through an IEDN virtual server if necessary.

**Example:** Router on Virtual Server that belongs to two VLANs and implements IP Filtering (Firewall)

Virtual Server Operating Systems should not forward from one VLAN ID to another within the IEDN via Layer 3 routing since this can conflict with the Layer 2 VLAN security that the administrator handling the Network Virtualization Manager tasks has established. By setting up routing within a Virtual Server you now have another administrator potentially interfering with the security that was defined at the HMC. In fact, you have seen that on z/OS IP Forwarding has been disabled when traffic comes in on an OSX port and wishes to exit on the same or a different OSX port.

If the implementer of the Virtual Server inside the zBX decides to permit routing through that virtual server, then this visual illustrates how the packets would flow: from VLAN10 at TCPIP3, over the OSX port, through the IEDN to the TOR, and then to VS-1011B on VLAN10; VS-1011B inspects the packet with IP Filtering or an internal Firewall and if allowed, routes the packet to VLAN11 and out over the IEDN to the Virtual Server11D which is connected to VLAN11. Remember that this routing should be permitted ONLY IF IP FILTERING has been implemented to permit the traffic to flow as depicted. Bear in mind that currently there are no stateful IP Filtering or Firewall mechanisms that can run on a blade or an LPAR.
Security in the Ensemble

August 2011: The z196 server has achieved a Common Criteria certification at an EAL5+ level. The certification is now listed on BSI’s website https://www.bsi.bund.de/ContentBSI/EN/Topics/Certification/newcertificates.html
The IBM Security Framework provides a model for selecting, designing, and monitoring technologies to protect all aspects of an IT organization.

IBM provides the professional services to assess an organization's needs for security with regard to compliance mandates and general security requirements. These services can design, implement, and manage security technologies and can recommend hardware and software solutions for an organization.

**Security Services and Mechanisms**

<table>
<thead>
<tr>
<th>MANAGEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Security Services and Mechanisms</strong></td>
</tr>
<tr>
<td><strong>Security Services and Mechanisms</strong></td>
</tr>
<tr>
<td><strong>Security Services and Mechanisms</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authentication</th>
<th>Access Control</th>
<th>Confidentiality</th>
<th>Data Integrity</th>
<th>Non-Repudiation</th>
<th>Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logon IDs</td>
<td>Access Control Lists</td>
<td>Encryption (based on Selected Algorithms, e.g. 3DES, AES, etc.)</td>
<td>Checksum Message Integrity code</td>
<td>► Origin</td>
<td></td>
</tr>
<tr>
<td>Passwords</td>
<td>Security Labels</td>
<td>Data masking</td>
<td>Digital Signatures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pass Tickets</td>
<td>Roles</td>
<td>Encryption (based on Selected Algorithms, e.g. 3DES, AES, etc.)</td>
<td>Digital Certificates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Certificates</td>
<td>Physical Barriers</td>
<td>Data masking</td>
<td>Trusted Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private Keys</td>
<td></td>
<td>Encryption (based on Selected Algorithms, e.g. 3DES, AES, etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart Cards and PINS</td>
<td></td>
<td>Checksum Message Integrity code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCMCIA Cards</td>
<td></td>
<td>Digital Signatures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biometrics</td>
<td></td>
<td>Digital Certificates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trusted Time</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This is an older version of the ISO security model. Note the entry for "Governance" and "Logging." This is not part of the ISO model, but it is nevertheless integral for any security implementation. We have added it here to show its importance.

You have now seen a couple of architectures for security. What is the difference?

The IBM Security Framework describes WHAT needs to be protected in an IP installation.

This ISO Security architecture describes:

- **HOW** to protect the data (i.e., which Services should be used to protect data and resources)
- **WHICH** ways to protect the data (i.e., which security mechanisms could be used to protect the data and resources)

Ensemble networking provides many layers of security to ensure that data is not compromised. The next visuals outline for you many of these methods.

**NOTE:** We believe there is no reason to implement IPSec for encryption because we believe that encryption should not be required in the IEDN. However, if the customer wants to encrypt, we would recommend SSL/TLS/AT-TLS.
Frequently Asked Questions about Securing the Ensemble Network

• **How do I limit and secure access to the Ensemble Definitions?**
  – Use controls at the HMC (Unified Resource Manager functions)
  – Use existing HMC controls
    • Reference the HMC White Paper (see Appendix)

• **How do I limit and secure access to the Ensemble Virtual Servers and Networks?**
  – Authorize the Virtual Servers to become Ensemble Members
  – Authorize the Virtual Servers to send data across the Ensemble Networks
  – Authorize the Virtual Servers to exploit only certain VLAN IDs
  – Use existing security techniques
    • UserID and passwords
    • Access controls to access storage
    • Firewalls and IP Filtering
    • Etc.

• **How do I limit access to the Ensemble by the External Networks?**
  – Deploy routers and Firewalls to Permit or Deny traffic
  – Implement controls at the Unified Resource Manager and at the LPARs to limit access
Security Enforcement and Filtering in the Ensemble

- **Deploy traditional security measures -- although many are unnecessary in an Ensemble**
- **Unified Resource Manager Provides Centralized Control of the Enforcement Definitions in the IEDN**
  - VMACs (Physical Adapter Addresses)
  - VLANs (Virtual Local Area Network IDs)
  - IP Addresses
- **Hypervisors in the Ensemble ENFORCE the Definitions**
  - VMACs
  - VLANs
  - IP Addresses
- **The Top of Rack Switch FILTERs for ...**
  - Correct external MAC Addresses (Physical Addresses)
  - Correct external VLAN IDs
- **An Ensemble Member in the z196 LPARs PERMITS or DENIES**
  - Routing into the IEDN
  - ACCESS Controls into the IEDN

Traditional security measures can continue to be deployed in the Ensemble. However, the enhanced security of the Ensemble is provided by new features, definitions, and enforcement methods in the Ensemble. The Unified Resource Manager, the Hypervisors, and the Top of Rack switches provide the means to control the enhanced security.

Specifically, enhanced Ensemble Security is defined in the Unified Resource Manager panels. The definitions are pushed to the Hypervisors of the Virtual Servers in the Ensemble and enforced there. In addition, the Top of Rack switches are also responsible for implementing certain security controls, like VMAC filtering and VLAN enforcement.

Finally, Ensemble Operating System controls can provide additional security for entering the IEDN.

What is a Hypervisor: An operating system or piece of code that allows you to run other operating systems on top of it. In the zEnterprise Ensemble, the following components constitute Hypervisors:

- **PR/SM (z196)**
- **OSX OSA Ports** implement a version of a VSwitch in order to support multiple LPARs or Virtual Servers that can share the OSA Port; the internal VSwitch of the OSX port is a Hypervisor component for PR/SM
- **z/VM**, where a VSwitch is a Hypervisor component (z/VM)
- **pHyp or PowerVM**, where VIOS and a VBridge or VSwitch are components of the Hypervisor (AIX on POWER Series)
- **KVM (or xHYP)**, which is a HyperVisor that we expect to see for x-Blades in First Half of 2011.
Role-Based Access Controls with the Unified Resource Manager

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensemble Administrator</td>
<td>Responsible for creating and managing the zEnterprise ensemble Create Ensemble, Add Member…</td>
</tr>
<tr>
<td>Virtual Network Administrator</td>
<td>Responsible for Managing Virtual Networks, Hosts, and MAC Prefixes Manage Virtual Networks, Add Hosts to Virtual Networks, Create VLAN IDs…</td>
</tr>
<tr>
<td>Virtual Server Administrator</td>
<td>Responsible for managing virtual servers New/Modify Virtual Server, Add Virtual Disk, Migrate…</td>
</tr>
<tr>
<td>Virtual Server Operator</td>
<td>Responsible for performing and scheduling virtual server activation/deactivation, mounting virtual media Activate, Deactivate, Mount Virtual Media, Console session…</td>
</tr>
<tr>
<td>Storage Resource Administrator</td>
<td>Responsible for managing storage resources – Storage Access Lists, WWPNs, z/VM Storage Groups Export WWPN, Import SAL, Add Storage Resources…</td>
</tr>
<tr>
<td>Workload Administrator</td>
<td>Responsible for managing workloads New/Modify workload, Add/Remove Virtual Servers…</td>
</tr>
<tr>
<td>Performance Management Administrator</td>
<td>Responsible for managing performance policies New/Modify performance policy, Import policy</td>
</tr>
<tr>
<td>Performance Management Operator</td>
<td>Responsible for performing and scheduling policy activations and creating threshold notifications Activate, Export Policy, Monitor System Events</td>
</tr>
<tr>
<td>Energy Management Administrator</td>
<td>Responsible for managing power settings including power capping and power savings Set Power Cap, Set Power Savings Mode, Set zBX Power Policy</td>
</tr>
</tbody>
</table>

- New task and resource roles enable isolation across management disciplines
- New predefined users EnsOperator and EnsAdmin

The role-based access control in the Unified Resource Manager provides granularity for the different functions that are used to manage an Ensemble.

Since these panels are available only from the HMC, the security for the Ensemble functions is another security layer already available in the HMC itself.

**Security in the HMC and the Unified Resource Manager:**

The HMC continues to enjoy the security controls inherent in the System z Console. Please see IBM System z Hardware Management Console Security White Paper by Kurt Schroeder (schroedk@us.ibm.com) published in Sept. 2008 and available at:


This paper explains: “The HMC Licensed Internal Code includes a full-function firewall that is used to control network access to the HMC. As previously described, by default the HMC allows for virtually no inbound network traffic. As different features of the HMC are enabled (e.g., remote access, SNMP based automation etc.) additional inbound network traffic is allowed.”

In addition, the Unified Resource Manager screens of the console enjoy additional security controls, like role-based security and operations over the closed, Intranode Management network (INMN).
### INMN:

1. **Authentication and Access Control:**
   - 1. Isolated and non-configurable VLANs in the INMN
     - Exploited only by Management Applications communicating between the Support Element and the Virtual Servers
   - 2. Virtual Servers cannot communicate with each other over the INMN
   - 3. **ISOLATE** over a shared OSM OSA is enforced.
   - 4. Restrict access to OSM with SERVAUTH Class
     - **EZB.OSM.sysname.tcpname** (z/OS only)
   - 5. IP Filtering is optional
     - Can implement Security Classes for IP Filtering

2. **External Network Access impossible:** the INMN ends at the HMC and Support Element.
IEDN:

1. Reduce the Scope of Security Vulnerability through Elimination of Routing Hops and Network Administration
2. Authentication and Access Control:
   1. Only authorized servers can access the IEDN
      • **OSDSIM** support in z/VM grants access to certain Guests on the VSwitch under VM
   2. Isolated VLANs in the IEDN with VLAN Enforcement for nodes on the VLAN.
   3. By default: No routing within the IEDN between different VLAN IDs
      • VMAC enforcement is required within the IEDN; VMAC(MAC) Filtering is recommended between External Customer Network and the IEDN
      • **ISOLATE** over a shared OSX OSA is optional, but plan carefully!
      • If necessary, implement Firewall for routing between VLANs in the IEDN
   4. Restrict Access with z/OS TCP/IP **NETACCESS** Controls
   5. IP Filtering (IPSec) is optional
      • z/OS can implement Security Classes for IP Filtering (IPSec)
3. **Confidentiality** (Encryption) is optional.
4. **External Network Access** permitted only if connection is authorized in the LPAR or the TOR
   1. Recommendation: Implement Customer Firewalls and/or IP Filtering
   2. Implement Multi-Level security where desirable.

Some topics mentioned here have not been part of the main body of this presentation. Please review the materials in the Appendix and the References at the back of this document for more information.
Review: Connecting the Customer Data Network to the intraensemble Data Network

1. Enter through an OSD Connection attached to an Ensemble Member.

2. Enter through a Router connection to the TOR.

You have already seen this slide. We now review it.

If you decide to permit communication between the External Customer Data Network and the Ensemble, you can keep this path secure.

Note that for security purposes a Layer 2 connection from the external network into the TOR is not supported – the connection must be using Layer 3 protocols. (Bridge Protocol Data Units – BPDUs – at Layer 2 cannot be successfully exchanged between an external Layer 2 switch and the IEDN TOR. Spanning Tree Protocol (STP) messages that might be received from external switches are filtered out at the TOR. This together with an external Firewall is to protect the security of the IEDN network by avoiding VLAN ID collisions. For example, if a customer were to attach an external switch to the TOR, and BPDUs and STP messages were not being filtered out, a customer’s external VLAN ID might be the same VLAN ID used within the IEDN and thus mistakenly cause the interconnection of external VLAN segments to the IEDN VLAN segments, thus impinging the security of the IEDN.)

First, determine whether you want to do this. Second, determine how you will secure this connection. Be aware of the fact that the TOR performs VLAN ID enforcement for connections to servers outside the zBX that are not attached to an OSX OSA port. The VLAN ID enforcement for any Virtual Server attached to an OSX works as follows: If a z/OS Native LPAR is on the OSX, then the OSX performs the VLAN ID enforcement; If the Virtual Server is under z/VM and attached to a VSwitch, then the VSwitch performs the VLAN ID enforcement. If an ISAOPT appliance is on the zBX, then the TOR performs the VLAN ID enforcement. The hypervisors on the Blade of the Virtual Servers of the zBX perform VLAN ID Enforcement in the zBX. (You configure the authorized VLAN IDs at the HMC as part of the Network Virtualization infrastructure.)

Remember that Security protection is much more than just inserting a firewall along a path. It encompasses all layers of the IP Stack: Application Security Mechanisms (Access Control Lists, Userid and Password checking, mapping mechanisms), Transport Security Mechanisms (SSL/TLS, AT-TLS), IP Layer Security Mechanisms (IPSec, IP Filtering, Intrusion Detection Services, Network Address Tables), Data Link Control Security Mechanisms (MAC Address Filtering, VLAN Segmentation or Segregation), and many more mechanisms too numerous to mention here.

With regard to MAC Filtering, the TOR can perform MAC filtering for external MAC Addresses based on Unified Resource Manager definitions. The MACs within the IEDN are managed by the Network Virtualization Manager. All MACs are allowed that originate from within the IEDN (they are managed by NVM). z/VM VSwitch MAC Protect function for Layer 2 is on by default for IEDN type VSwitches. This MAC Protect function enforces that a VMAC sent during guest link initialization (SETVMAC) matches with what has been assigned by the zVM hypervisor. In addition, all SOURCE MAC addresses on egress frames from the guest are verified to insure that only the assigned VMAC for the guest is being sent on outbound data transfers. This eliminates any attempt by the guest to spoof its source MAC address.

Note on use of VMACs: When an Operating System on z is using layer 2, it performs an ARP and builds Ethernet headers with VMAC. z/OS does not use layer 2 and so ARP is handled by the OSA and the OSA builds the Ethernet header. On a VSwitch under z/VM, you might have a Linux guest using Layer 2. In this case the VSwitch builds the Frame Header with the VMAC in it. But, if the guest system is using Layer 3 on a VSwitch, then a frame header is not necessary and the packet is forwarded over the VSwitch using only the IP address.

If you are still interested in introducing firewalls consider these possibilities:

Firewalls
- IP Filtering in z/OS Policy Agent (not stateful)
- IP Filtering in z/OS Policy Agent (not stateful)
- IP Filtering in z/OS Policy Agent (not stateful)
- Other IP Filtering mechanisms (could be stateful or not)
- Firewall in front of LPAR that is attached to an OSD OSA
- Firewall in front of TOR in External network
SERVAUTH Classes: NETACCESS CONTROLS for IEDN
MAC ADDRESS FILTERING at the TOR
MultiLevel Security
What May Catch You by Surprise: Simplicity of IPv6 & Security

- IPv6 for INMN need not be a “Big Deal"
  - You need not learn IPv6 (invisible outside the INMN)
  - Be aware of Display Output Changes when a z/OS stack is IPv6-enabled
  - IPv6 is completely transparent in z/VM
- INMN Implementation is Simple and Secure
  - Customer needs no control over the IPv6 link-local addresses, which are dynamically assigned
  - A CLOSED, FLAT network
  - Authorizations and OSM Access Control (z/OS) are required to reach it
  - IP Forwarding is disabled
- IEDN enablement adds only a few parameters
  - In z/OS Dynamic VTAM Definitions or Manual
  - In z/VM the Systems Management API builds the z/VM Directories and VSwitch definitions
- The IEDN network is inherently more secure than the External Customer Network
  - Customer decides if External Connectivity is needed
  - It is an Internal Network secured by
    - Additional Security Definitions
    - VLAN Routing with VLAN Enforcement
    - VMAC Enforcement
    - VMAC Protect Function in z/VM VSwitch and in Network Virtualization Mgr.
    - VMAC Filtering via HMC
    - IEDN IP Forwarding is disabled on z/OS
  - Secured by traditional means to secure all security layers, but outside the centralized trust and access controls of the Unified Resource Manager definitions.
  - Secured by traditional means to secure the network layer, but outside the centralized trust and access controls of the Unified Resource Manager definitions.
  - Additional definitions in the HMC authorized only to specific administrators
  - VLAN Routing with VLAN Enforcement
  - z/VM VMAC Protect Function enforces VMACs assigned and generated with Ensemble Management
  - VMAC Filtering via HMC
  - Multilevel Security (z/OS)
  - Secured by traditional networking security:
    - Firewall Filtering (not stateful) that can be loaded into the Operating Systems
    - External stateful firewall filtering
    - Network Access Controls (z/OS)
    - Technologies to protect Data in Transit (IPSec, SSL/TLS, etc.)
- The Virtual Servers, including z/OS, z/VM, and Linux on z become Dual-Mode Stacks
  - BPXPRMxx
  - In z/OS there are changes for UNIX initialization
  - IPv4 and IPv6 interfaces and routing
  - In z/OS IPv6 is “INTERFACE” only (not DEVICE/LINK)
  - On z/OS: Long Format only of NETSTAT
  - NETSTAT ROUTE for IPv6 and not NETSTAT GATE
- INMN Implementation is Simple and Secure
  - The INMN enablement results in “Stateless Autoconfiguration” of IPv6 addresses
  - Customer needs no control over the IPv6 link-local addresses, which are dynamically assigned
  - The INMN network need not be secured beyond what you would normally do with access to the HMC
  - A CLOSED, FLAT network
  - Authorizations and OSM Access Control (z/OS) are required to reach it
- IEDN enablement adds only a few parameters to definitions
  - allows customer to define IPv4 addresses as usual, or ...
  - IPv6 addresses
  - Dynamically, through stateless autoconfiguration (advertisements from routers build the addresses)
  - Manually, through customer definition.
- The IEDN network is inherently more secure than the External Customer Network
  - It is an Internal Network secured by
    - Additional definitions in the HMC authorized only to specific administrators
    - VLAN Routing with VLAN Enforcement
    - z/VM VMAC Protect Function enforces VMACs assigned and generated with Ensemble Management
    - VMAC Filtering via HMC
  - Additional definitions in RACF to limit access
  - Secured by all other means already available to the Operating Systems loaded into the Virtual Servers – although these security measures are at many points of control and executed by many different entities instead of centrally at the HMC with Unified Resource Manager firmware. That is, these non-IEDN methods are subject to errors and trust issues when the number of administrators allowed to make decisions and definitions concerning security broadens the scope of security control.
    - Multilevel Security (z/OS)
  - Secured by traditional networking security:
    - Firewall Filtering (not stateful) that can be loaded into the Operating Systems
    - External stateful firewall filtering
    - Network Access Controls (z/OS)
    - Technologies to protect Data in Transit

What May Catch You by Surprise: Non-Disruptive Test Bed for IPv6

- Securely Isolate an IPv6 Test (or even Production) Network from the IPv4 network.

The visual shows you that you could isolate certain Virtual Servers in the IEDN and then use them for testing IPv6 network connectivity.
Appendices: Additional Networking Topics
Another Topology for the Ensemble:
Sharing a zBX
In the visual you see two zEnterprise nodes. They are sharing a single zBX. Node 1 manages the zBX and Node 2 is connected to the zBX only over the IEDN. (Contrast this image with one depicted earlier in this presentation where two z196 nodes, each with its own zBX, participate in the Ensemble.)

**HMC Communications:**
The HMC communicates with each node over the Customer Management Network by connecting through the Bulk Power Hub (BPH) switch which is attached to both the Support Element (SE) and the INMN of each node. The same HMC is used to manage all nodes within the Ensemble using the same network connectivity.

**INMN Communications:**
The intranode management network interconnects only within a node; it does not span nodes. Therefore you see that there is an INMN within Node 1, where the z196 manages or “owns” the zBX. There is a second INMN within Node 2, where the only member of Node 2 is the z196 itself.

**IEDN Communications:**
The intraensemble data network spans nodes in the ensemble. Therefore you see a single IEDN interconnecting Node 1 and Node 2.

**Advantages of such a configuration:** The two z196 CECs are tied together in a flat network that is under the control of the security and administrative tasks of the Unified Resource Manager. This is a one-hop network interconnecting the two CECs by means of the 10Gig Top of Rack Switches. With this configuration, latency over the network is reduced, routing overhead is reduced, and encryption is not required.
IP Address Registration in the OSA
### OSA Address Table (OAT) Registration Details for z/OS Single Connection per Stack per Subnet (2)

<table>
<thead>
<tr>
<th>Connection Definition Type</th>
<th>Subnet coded?</th>
<th>Basic definition (no VMACs)</th>
<th>VMAC ROUTEALL configured?</th>
<th>VMAC ROUTECL configured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEVICE and LINK (3)</td>
<td>N/A</td>
<td>All active IP addresses in HOME List (1) - OSA (3) - CTC (2) - XCF - HyperSockets, etc. - All VIPAs</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>INTERFACE (2)</td>
<td>No</td>
<td>All active IP addresses in HOME List (1) - OSA (3) - CTC (2) - XCF - HyperSockets, etc. - All VIPAs</td>
<td>OSA IP address(2) - All VIPAs (4)</td>
<td>-</td>
</tr>
<tr>
<td>INTERFACE (2)</td>
<td>Yes</td>
<td>Most active IP addresses in HOME List (1) - OSA (3) - CTC (2) - XCF - HyperSockets, etc. - VIPAs in same subnet as OSA (4) - All other VIPAs (5)</td>
<td>OSA IP address(2) - All VIPAs in same subnet as OSA (4)</td>
<td>-</td>
</tr>
</tbody>
</table>

**GENERAL NOTES:**
- z/OS IP Address OSA Registration for:
  1. Inbound routing and
  2. ARP offload
- OSA/SF "GET OAT" and VLANID influences routing
  1. Addresses registered for ARP offload purposes
- For VLANID purposes:
  + Consider IP address for VMAC ROUTECL
  + Ignore IP address for VMAC ROUTEALL

This chart is meant to help you understand how the IP addresses you choose in the IEDN may be located or not through the OSX OSA Ports on the LPARs in the IEDN. The chart is specific to z/OS and any operating system may choose or not to register certain addresses. Generally speaking, if a z/VM VSwitch is told by an operating system guest to register an IP address, that IP address will be registered in the OAT.

**General NOTES:**
- z/OS registers IP addresses for two purposes: for inbound routing and for ARP offload (responding to ARP requests). The OSA/SF "GET OAT" function displays only addresses that have been registered for ARP offload purposes. The coding of VLANID does not influence ARP registration; it does, however, influence routing; if VLAN is coded on the z/OS or z/VM stack, only correctly tagged packets are considered for routing, after which the IP address will be considered for VMAC ROUTECL and ignored for VMAC ROUTEALL.
More Information on IPv6
IPv6 Benefits: An Overview

IPv4 Address: 32 bits (4 octets)
Example: 192.168.24.100

IPv6 Address: 128 bits long
Link-local Address example: fe80:0:0:0:6:2900:46DC:217C
Global Address example: 50c9:c2d5:0:0:9:67:115:5

- A dramatically larger address space,
  -- said to be sufficient for the next 30 years.
- Globally unique and hierarchical addressing,
  -- based on prefixes rather than on address classes to keep routing tables small and backbone routing efficient.
- Multicasting instead of broadcasting.
- Class of service to distinguish between different types of traffic.
- A built-in mechanism for autoconfiguration of network interfaces.
- Built-in authentication and encryption.
- Encapsulation of itself and other protocols.
- Transition methods to migrate from IPv4.
- Compatibility methods to coexist and communicate with IPv4.

The first thing that most people think of when they hear about IPv6 is that the IP address is much longer (128 bits) than an IPv4 address (32 bits). IPv4 addresses are described in RFC1166: “Internet Numbers.” IPv4 addresses are represented in dotted-decimal format. The 32-bit address is divided along 8-bit boundaries. Each set of 8 bits is converted to its decimal equivalent and separated by periods. Each IP address consist of an IP network id and an IP host id on that IP network.

In contrast, IPv6 addresses are 128 bits divided along 16-bit boundaries. Therefore, IPv6 notation is eight 16 bit integers separated by colons. Each 16-bit block is converted to a 4-digit hexadecimal number -- still separated by colons. One group of multiple zeroes can be represented with a double colon. Leading zeroes within each individual field can be omitted. The resulting representation is called colon-hexadecimal. This notation is described in RFC 3513 (obsoletes RFC 2373): “IP Version 6 Addressing Architecture.” An IPv6 address prefix is analogous to the IPv4 network or subnetwork that should be applied to the complete IPv6 address. IPv6 addresses have different types of prefixes: a link-local prefix for addresses that are not routable outside of the current link; a global prefix for addresses that are routable outside of the current link.

This visual also shows you other benefits of IPv6.
IPv6 Unicast Addresses (!) and Scopes

<table>
<thead>
<tr>
<th>Scope Prefix</th>
<th>Format Prefix</th>
<th>Scope Prefix</th>
<th>Format Prefix</th>
<th>Scope Prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast</td>
<td>0</td>
<td>9</td>
<td>64</td>
<td>127</td>
</tr>
<tr>
<td>Link-local Scope</td>
<td>10 bits</td>
<td>54 bits</td>
<td>64 bits Interface ID</td>
<td></td>
</tr>
<tr>
<td>Unique local Scope (ULA)</td>
<td>7 bits</td>
<td>40 bits</td>
<td>16 bits</td>
<td>64 bits Interface ID</td>
</tr>
<tr>
<td>Global Scope</td>
<td>RFC 3513</td>
<td>3 bits</td>
<td>61 bits</td>
<td>64 bits Interface ID</td>
</tr>
<tr>
<td></td>
<td></td>
<td>anything else</td>
<td>variable &quot;subnet&quot;</td>
<td>MAC, Other Interface ID</td>
</tr>
</tbody>
</table>

IPv6 prefixes are similar to IPv4 subnets. That is, an IPv6 prefix route will route to all the host addresses that fall under the prefix. IPv6 prefix routes are specified with XXX:: in IPv4 prefixes are specified. For IPv6 routing to work in IPv4 like IPv6, the addresses all have to be in the same prefix just as they must be in the same subnet for IPv4. The rest of these notes explains the layout of IPv6 prefixes and addresses.

The left-most bits of the addresses are divided into two portions:
- the "format prefix" defined in RFC 3513 (obsoletes RFC 2373); the globally unique IP addresses assigned to a company have format prefixes that are called "global routing prefixes" --- and the "network" or "subnet" portion -- which, when combined with the first left-most bits is also called a "prefix." You can see the confusion that arises with the terminology already!!
- The "format prefix" or "scope prefix" indicates the type of address or "scope" of address that is being represented. IPv6 RFCs define several different "scopes" or categories of standard unicast addresses: Link-Local, Unique Local, Global.
  - RFC2373 indicates that the globally aggregatable scope is identified by "001" in the first three bits; RFC 3513 indicates that "everything else" other than the UNSPECIFIED, LOOPBACK, Link-Local, and Site-Local prefixes are Global scope. Addresses with a Global Scope are also called "Globally Aggregatable Addresses" because the routing in IPv6 allows the addresses to be aggregated into small routing trees using the CIDR concept.

From Wikipedia: In October 2005, RFC 4193 was published, reserving the address block fc00::/7 for use in private IPv6 networks, and defining the associated term unique local addresses. The address block fc00::/7 is divided into two /8 groups:
- The block fc00::/8 has not been defined yet. It has been proposed to be managed by an allocation authority, but this has not gained acceptance in the IETF.[1][2][3]

The range fd00::/8 is defined for 128 prefixes, formed by setting the 40 least-significant bits of the prefix to a randomly-generated bit string. This results in the format fd00:XXX:XXX:XXX: for address in this range. RFC 4193 offers a suggestion for generating the random identifier to obtain a minimum-quality result if the user does not have access to a good source of random numbers. (See following for sample random number generation methods: [https://www.ultratools.com/tools/rangeGenerator](https://www.ultratools.com/tools/rangeGenerator) and [http://bitace.com/ipv6calc/](http://bitace.com/ipv6calc/))

Unique Local Unicast Addresses (ULAs) are recommended for the test environment within an installation. Unlike site-local addresses (deprecated in 2004), a site may have more than one of these prefixes and use them at the same time.

Again, the left-most bits that identify these scopes are often called "Scope or Format Prefixes," but the term "Prefix" is also applied to everything preceding the right-most 64-bit Interface ID.

Note how the addresses in the visual all have in common the right-most 64-bit Interface ID. This portion of the address is analogous to the "host ID" portion of the IPv4 address format. This identifies a unique adapter on an internet node. The interface ID need not be the hardware address of an interface. It can be manually configured or dynamically configured. RFC 3041, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6," defines two types of interface designations for the Interface ID:
- Unique Stable IP Addresses (Manually configured, configured via DHCP server, autoconfigured with factory setting (e.g., MAC address));
- Temporary Transient IP Addresses (Dynamically configured with a Random Number)

Each of these address scopes has multiple variants, revolving around UNICAST and MULTICAST options. This page shows you the address layouts for UNICAST addresses.

Unicast addresses identify a single interface. A packet sent to a unicast address is delivered to the interface identified by that address. This is the same concept with which you are already familiar in IPv4.

Anycast addresses identify a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the "nearest" one, according to the routing protocols' measure of distance). This is a concept not used in IPv4.

Multicast addresses identify a set of interfaces (typically belonging to different nodes). A packet sent to a multicast address is delivered to all interfaces identified by that address. This is the same concept with which you are familiar in IPv4. Routing protocols like RIP and OSPF use multicast addresses, but so can other applications.

There are no broadcast addresses in IPv6, their function being superseded by multicast addresses.
Special Unicast Address: Link-local

- **Link-local Scope / Address - unique on a link only**
  - "Format prefix" or "Scope Prefix" is FE80::/10
  - Used for addressing between stations on the same link or LAN
  - **Cannot be routed**
    - One link-local required for each interface (excluding VIPA and loopback)
    - z/OS CS only allows a single link-local address per interface.

Communication using Link-Local Addresses

- Autoconfigured
- Dynamically by stack
- Neighbor discovery
- Networks without routers

The scope of a packet's source and destination addresses controls where in the network the packet will be routed. Every IPv6 interface will have a link-local address, which is automatically generated by the stack (considered one type of autoconfiguration).

The link-local "format prefix" -- now called a "link-local routing prefix" or "link-local scope prefix" -- is FE80::.

Unique Local Unicast Addresses are manually configured.

Site-local (now deprecated) and global address can either be manually configured or dynamically generated.

A packet with a link-local source/destination address will not leave its originating LAN. A router receiving the packet will not forward it onto another physical LAN.

Used for any kind of temporary network
- Dynamically configured by stack.
- Neighbor discovery
- Networks without routers

So remember: A packet with a link-local source/destination address will not leave its originating LAN. A router receiving the packet will not forward it onto another physical LAN.

Used for any kind of temporary network: Autoconfiguration, Neighbor discovery, Networks without routers
IPv6 Concepts: Global Scope

Special Unicast Addresses: Global

- **Global Scope / Addresses** - unique within Internet or Extranet.
  - *Format prefix* or *Scope Prefix* for Unicast typically starts at 200n
  - Must be used by routers in the global internet

By now you know that you need not know much about IPv6 at all, since the only IPv6 requirement for Ensemble networking is simply that the Operating System loaded into the Virtual Server must be enabled for IPv6. This does not mean that you need to assign IPv6 addresses or learn how to define IPv6 interfaces or how to route IPv6 packets. The IPv6 requirement of Ensemble Networking only requires that you anticipate the few changes that IPv6 enablement causes in your Operating System definitions and in certain command displays.

Although the IPv6 requirement need not be a major concern for you at all, we nevertheless want to show you the type of IPv6 address that you might implement in your Ensemble if you choose to use IPv6 over the IEDN network. If you decide to implement IPv6 fully, you will probably be implementing Global Unicast Addresses, as depicted on this visual. (Our remaining examples for Ensemble networking show addressing and routing with IPv4 addressing.)

The scope of a packet's source and destination addresses controls where in the network the packet will be routed. Remember: The scope of a packet's source and destination addresses controls where in the network the packet will be routed.

So how do you communicate across the internet or to any business partners who are not part of your site? You need to use addresses with a GLOBAL scope. A packet with global source/destination address can be routed anywhere.

Can be autoconfigured using Router Advertisements (Stateless autoconfiguration) or using DHCPv6 (stateful autoconfiguration (not supported by CS for z/OS V1R4), or...

Can be manually configured through IP Stack configuration.

To communicate across LANs or physical networks, you may use site local addresses. A packet with a site local source/destination address will not leave its originating site.

A packet with global source/destination address can be routed anywhere.

If a packet cannot be forwarded due to reaching a scope boundary, an ICMPv6 error message is returned stating BEYOND SCOPE.

The "Format prefix" or "Scope Prefix" for Globally Routable Unicast addresses typically starts at 200n, but it can be anything else not already defined by the RFC 3513.

**Special Unicast Addresses: Global**

**Global Scope / Addresses** - unique within Internet or Extranet.

"Format prefix" or "Scope Prefix" for Unicast typically starts at 200n

Must be used by routers in the global internet

Global @, 2nnn or 3nnn or other, H2, R4, R5, R1, R2, R3, H1

Communication using Addresses with a Global Scope:

- H5& to H3%
- H5& to H4#
- H5& to H1=
- H5& to H2$

Internet, Extranet; H5; &; %; #; =; $
### IPv4 vs. IPv6 Characteristics

<table>
<thead>
<tr>
<th>Feature</th>
<th>IPv4</th>
<th>IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Addressing</strong></td>
<td>32 bits (4 bytes) -- &gt; 4,200,000,000 addresses</td>
<td>128 bits (16 bytes) -- 340 (billion) addresses</td>
</tr>
<tr>
<td><strong>Communicating to all on a subnet</strong></td>
<td>Broadcast Addresses</td>
<td>Scoped Multicast Addresses</td>
</tr>
<tr>
<td><strong>Fragmentation</strong></td>
<td>Supported at originating and intermediate nodes</td>
<td>Supported only at originating nodes</td>
</tr>
<tr>
<td><strong>Checksum</strong></td>
<td>Included in IP Header</td>
<td>Not included in IP Header</td>
</tr>
<tr>
<td><strong>IPSec</strong></td>
<td>Optional</td>
<td>Required for full IPv6</td>
</tr>
<tr>
<td><strong>Discovery of best default gateway</strong></td>
<td>Optional; with ICMP Route Discovery</td>
<td>Required: ICMPv6 Router Solicitation and Router Advertisement</td>
</tr>
<tr>
<td><strong>Resolving IP layer address to link layer address</strong></td>
<td>ARP (Address Resolution Protocol)</td>
<td>Multicast Neighbor Solicitation Messages</td>
</tr>
<tr>
<td><strong>Local Subnet Group Membership</strong></td>
<td>Internet Group Management Protocol (IGMP)</td>
<td>Multicast Listener Discovery (MLD)</td>
</tr>
<tr>
<td><strong>Address Configuration</strong></td>
<td>Manually or through DHCP</td>
<td>Automatically assigned via stateless address configuration; or via DHCPv6, or manually</td>
</tr>
<tr>
<td><strong>DNS Configuration</strong></td>
<td>Use &quot;A&quot; records for host name/address mapping</td>
<td>Use &quot;AAAA&quot; or &quot;A6&quot; records for mapping</td>
</tr>
<tr>
<td></td>
<td>Use &quot;PTR&quot; records in IN-ADDR.ARPA domain for address/name mapping</td>
<td>Use &quot;PTR&quot; records in IP6.ARPA or IP6.INT domain for address/name mapping</td>
</tr>
<tr>
<td><strong>QoS Support</strong></td>
<td>Differentiated and Integrated Services</td>
<td>Differentiated and Integrated Services; also Flow Label for more granularity</td>
</tr>
<tr>
<td><strong>Payload Identification for QoS</strong></td>
<td>Not included in IP Header</td>
<td>Included in Flow Label</td>
</tr>
</tbody>
</table>

This chart represents a summary of the information present in Table 1.17 of the IPv6 Network and Application Design Guide Version 1 Release 4 (SC31-8885-00).

You have heard about most of these items during this topic; what you have not heard yet will be presented in subsequent lectures.

Additional differences:

- **IP Header Format**
  - IPv4 -- Variable: Min of 20 Bytes + Options
  - IPv6 -- 40 Bytes

- **IP Options**
  - IPv4 -- Part of IP Header
  - IPv6 -- Inserted as Extensions between IP Header and Payload

QoS, IPSec, DHCPv6, and Mobility are not part of the z/OS V1R4 Implementation of IPv6.
More Information on VLANs and IDs
A local area network (LAN) is a broadcast domain. Nodes on a LAN can communicate with each other without a router, and nodes on different LANs need a router to communicate. A virtual LAN (VLAN) is a configured logical grouping of nodes using switches. Nodes on a VLAN can communicate with each other as if they were on the same LAN, and nodes on different VLANs need a router to communicate. (Layer 3 routers can add, remove, or validate VLAN tags.) The IBM Open Systems Adapter provides support for IEEE standards 802.1q, which describes VLAN identifier tagging. (Note that currently the OSX implementation supports 802.1q only.) Deploying VLAN IDs allows a physical LAN to be partitioned or subdivided into discrete virtual LANs. This support is provided by the z/OS TCP/IP stack and the OSA-Express feature in QDIO mode. When you use VLAN IDs, the z/OS TCP/IP stack can have multiple connections to the same OSA-Express feature. One connection is allowed for each unique combination of VLAN ID and IP version (IPv4 or IPv6).

Note in the top half of the visual how one network takes advantage of the physical connectivity. In the bottom half of the visual, we have split the physical LAN into two VLANs: one with VLAN ID of 10 and another with VLAN ID of 11. Z/OS and z/VM are implemented with a VLAN technology called “Global VLAN.” With Global VLAN, z/OS and z/VM can define a VLAN ID which is then registered in the OSA port. The OSA port then performs the VLAN tagging. The implementation of Global VLAN causes the stacks to be technically unaware of the VLAN, or “vlan-unaware.” However, many people find this subtle distinction confusing and refer to z/OS and zVM as “vlan-aware” stacks since they can define a VLAN ID for a LAN connection. For Linux on z (native), the TCP/IP stack itself performs the VLAN tagging, and, thus, Linux on z when running native is not using the Global VLAN ID but rather the standard 802.1q implementation of VLAN. Linux on z is thus technically a “vlan-aware” stack. You may read about Global VLAN IDs at:


“A GLOBAL VLAN ID is OSA's VLAN support to provide access to a virtual LAN segment for a VLAN unaware host so the host can receive and send its network traffic. This host does not tag its outbound frames nor receive tagged inbound frames. The GLOBAL VLAN ID participates on the VLAN transparently with OSA handling all the tagging work (VLAN-unaware). A host device driver can register a Global VLAN ID with the OSA-Express adapter. Typically each host defines only one Global VLAN ID per connection. Some device drivers allow configuration of one VLAN ID for IPv4 and a second VLAN ID for IPv6. The OSA-Express will use the Global VLAN ID to tag frames and send out Gratuitous ARP requests (ARP requests to check for duplicate IP addresses) on behalf of the host. The NIC simulation in z/VM also provides this support, which is separate from the virtual switch support. The Global VLAN ID processing for the virtual NIC is performed prior to any virtual switch port ingress processing and after virtual switch port egress processing.

One example of this is in z/VM. You can specify the VLAN keyword on a LINK configuration statement for a QDIOETHERNET link to register a Global VLAN ID. In the past with OSD and the ability to code VLAN IDs on a Virtual Switch, the following recommendation has been made: “To reduce complexity and host TCP/IP configuration changes when configuring a virtual switch host connection, it is recommended that you do not configure a global VLAN ID for a host that will be connected to a trunk port. Instead, connect the host to an access port and authorize it for the desired VLAN ID. This assigns a port VLAN ID (pvid) for the access port and all VLAN operations occur within the virtual switch.” With Ensemble Networking, the z/VM VSwitch is the enforcement point for VLANs that have been assigned with the HMC through the Network Virtualization Management functions; in addition the TOR is coded in TRUNK Mode and not ACCESS Mode. Therefore, the z/VM guests that are capable of doing so should assign the VLAN ID that the HMC recognizes; then the VSwitch enforces the activation of that connection to the VSwitch using the appropriate VLAN ID.
VLAN Concepts: The VLAN ID and Layer 2 Forwarding

<table>
<thead>
<tr>
<th>Frame Header</th>
<th>IP Header</th>
<th>Transport Header</th>
<th>Data</th>
</tr>
</thead>
</table>

2. You can **forward** data packets using the VLAN ID contents of the Frame Header, or
3. You can **route** data packets using the contents of the IP Header.

### TCP/IP Stack Architectural Layers

<table>
<thead>
<tr>
<th>Layer</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Application Layer</td>
<td></td>
</tr>
<tr>
<td>4. Transport Layer</td>
<td>TCP, UDP Port Number</td>
</tr>
<tr>
<td>3. IP Routing Layer</td>
<td>IP Address</td>
</tr>
<tr>
<td>2. Data Link Control Layer</td>
<td><em>Physical Address (MAC Address)</em>&lt;br&gt;VLAN ID</td>
</tr>
<tr>
<td>1. Physical Layer</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

This chart is depicting the layout of a frame that carries the data to a node on a LAN. The data itself is packaged in an IP Packet. The IP Packet Header contains the target and the source IP Addresses. The Transport Header contains the port number of the applications that are to receive and send the data. However, to send the data across a network, the IP packet must be prefaced with a Frame Header which contains the Physical address – called the Medium Access Control Address (MAC Address), which is managed at Layer 2 of the TCP/IP stack. The LAN port that is represented by a MAC Address can further segment the data that is arriving at it by examining the VLAN Tag, which contains a VLAN ID. VLAN tagging is defined in the IETF standard 802.1q.

When using a VLAN ID to route data, this is called “Layer 2 Forwarding” or even sometimes “Layer 2 Routing.”

The IP routing table indicates how to route a packet by looking for the IP address of the next hop on the way to the final destination. The TCP/IP stack determines what MAC address is associated with the next-hop IP Address in order to build the frame header that is prefixed to the IP packet before it is sent out over the LAN.

Once the packet arrives at the Switch, the Switch needs to determine if there is a VLAN ID in the frame header over which it must send the message. A message that is destined for VLANx cannot be sent by a switch over VLANy. Assigning one VLAN ID to a set of nodes and a different VLAN ID to another set of nodes isolates the two sets of nodes so that they cannot intercept messages from each other over the LAN. Thus, assigning VLAN IDs provides a measure of security between the two types of nodes that are sharing a LAN segment. If a message from VLANx needs to arrive at VLANy, then a router – not a switch – is required to perform this transmission.

VLAN security has been called into question in recent years because a VLAN ID and a MAC address can be spoofed. You will see on subsequent charts that we don't believe these types of issues apply to our IEDN, since we control all access to the LAN via our hypervisors or OSX (i.e. an OS can't build its own LLC header). The exception is for the external ports on a TOR; the customer is responsible for securing this access.

**NOTE:** Technically, Layer 1, the Physical Layer, is assumed in the architectural descriptions of the TCP/IP Architecture. The description of TCP/IP in the RFCs and the literature usually talks about “4” layers, beginning with Layer 2. This explains why we have depicted the Physical Layer in square brackets: [1. Physical Layer].
This chart is depicting the layout of a frame that carries the data to a node on a LAN. The data itself is packaged in an IP Packet. The IP Packet Header contains the target and the source IP Addresses. The Transport Header contains the port number of the applications that are to receive and send the data. However, to send the data across a network, the IP packet must be prefaced with a Frame Header which contains the Physical address – called the Medium Access Control Address (MAC Address), which is managed at Layer 2 of the TCP/IP stack. The LAN port that is represented by a MAC Address can further segment the data that is arriving at it by examining the VLAN Tag, which contains a VLAN ID. VLAN tagging is defined in the IETF standard 802.1q. When using a VLAN ID to route data, this is called “Layer 2 Forwarding” or even sometimes “Layer 2 Routing.”

The IP routing table indicates how to route a packet by looking for the IP address of the next hop on the way to the final destination. The TCP/IP stack determines what MAC address is associated with the next-hop IP Address in order to build the frame header that is prefixed to the IP packet before it is sent out over the LAN. Once the packet arrives at the Switch, the Switch needs to determine if there is a VLAN ID in the frame header over which it must send the message. A message that is destined for VLANx cannot be sent by a switch over VLANy. Assigning one VLAN ID to a set of nodes and a different VLAN ID to another set of nodes isolates the two sets of nodes so that they cannot intercept messages from each other over the LAN. Thus, assigning VLAN IDs provides a measure of security between the two types of nodes that are sharing a LAN segment. If a message from VLANx needs to arrive at VLANy, then a router – not a switch – is required to perform this transmission.

VLAN security has been called into question in recent years because a VLAN ID and a MAC address can be spoofed. You will see on subsequent charts that we don’t believe these types of issues apply to our IEDN, since we control all access to the LAN via our hypervisors or OSX (i.e. an OS can't build its own LLC header). The exception is for the external ports on a TOR; the customer is responsible for securing this access.

NOTE: Technically, Layer 1, the Physical Layer, is assumed in the architectural descriptions of the TCP/IP Architecture. The description of TCP/IP in the RFCs and the literature usually talks about “4” layers, beginning with Layer 2. This explains why we have depicted the Physical Layer in square brackets: [1. Physical Layer].
Can Traffic “Hop” from one VLAN to Another? No, unless …

- A single VLAN Can Span Switches:

  ![Diagram showing VLAN spanning Switches]

- **Layer 2 Forwarding with a Switch – No Router Is Required for Layer 2 Routing.**

- One VLAN can be interconnected to another if a Router routes between the two VLAN IDs:

  ![Diagram showing VLAN interconnected with a Router]

- **Layer 3 Routing between VLAN 10 and VLAN 11 with a Router.**

A VLAN may have multiple physical hops that are interconnected using a SWITCH. The multiple physical segments of the VLAN can even be associated with completely different IP addresses but the SWITCH can connect the segments to provide a single path through the SWITCH. A SWITCH operates at what is called “LAYER 2” of the IP protocol stack, and examines the contents of the Frame Header to determine where to forward packets across physical segments in the switch.

If a connection path contains more than one VLAN ID, as you see in the second diagram above, you must use a ROUTER or a combination ROUTER/SWITCH to interconnect the parts of the entire path. A ROUTER operates at what is called “LAYER 3” of the IP protocol stack, and examines the contents of the IP Header to determine where to route (forward) the packets across physical segments of the path.
Even with a shared OSA port, VLAN IDs can prevent a message from reaching the wrong node attached to the same OSA Port. Note how a Virtual Mac (VMAC) has been assigned to each connection on the z/OS stack. Although the OSA itself has a Physical MAC, Ensemble Management provides a MAC prefix to the OSA, which then performs the VMAC assignment. (If this is not an ensemble scenario, then the coding in the stack provides the new VMAC.) With this assigned VMAC, the frame header will ensure that the message that is sent is properly sent to correct LPAR **ONLY IF** the VLAN IDs are also correct.

NOTE 1: z/OS implements what is called a GLOBAL VLAN ID as does z/VM.

NOTE 2: In z/OS, the OSA (due to ARP Offload) adds the tag if the Interface is IPv4 and identifies a VLAN ID on the Interface definition. The stack adds the tag for IPv6.

From the **z/OS Communications Server V1R12 IP Configuration Guide (SC31-8775):**

**VLAN switch concepts**

In conjunction with the IEEE standards, most VLAN aware switches recognize and support at least two modes, referred to as trunk and access modes. This support is provided on a switch port basis. The general concepts of the two modes are as follows:

**Trunk mode**

Indicates that the switch should allow all VLAN ID tagged packets to pass through the switch port without altering the VLAN ID. Trunk mode is intended for servers that are VLAN capable, and filters and processes all VLAN ID tagged packets. In trunk mode, the switch expects to see VLAN ID tagged packets inbound to the switch port.

**Access mode**

Indicates that the switch should filter on specific VLAN IDs and only allow packets that match the configured VLAN IDs to pass through the switch port. The VLAN ID is then removed from the packet before it is sent to the server (that is, VLAN ID filtering is controlled by the switch). In access mode, the switch expects to see packets without VLAN ID tags inbound to the switch port.
Details on Coding ISOLATE on shared OSD and OSX Ports in z/OS and z/VM
Secure Segregation of Traffic over Shared OSD: ISOLATE

BACKGROUND INFORMATION on ISOLATE:

ISOLATE was introduced in z/VM 5.3 with APAR VM64281 as “Port Isolation” on a VSwitch and then later in z/OS as “Connection Isolation” with the V1R11 release. Since VLAN enforcement did not exist on an OSD device the way it does on an OSX device, many customers were asking for a means to prohibit traffic forwarding across a shared OSA port.

VLANs, when properly implemented, can isolate traffic over a shared network and shared OSA port. The isolation is complete if all TCP/IP stacks that share an OSA port implement VLAN ID tagging and assign separate VLAN IDs.

Another method that is available to isolate traffic across a shared OSA port is by using “Connection Isolation” in either z/OS or z/VM. This method can be deployed for OSD CHPIDs with or without assigning a VLAN ID or a VMAC to the OSA port and simply requires that the parameter ISOLATE be coded on the QDIO Interface definition. If it is deemed necessary, you can also implement Connection Isolation over an OSX port, although the security implementations of Ensemble networking already perform isolation over the shared port when separate VLAN IDs are coded. Introducing ISOLATE without thought into the OSX environment can have unintended consequences as subsequent charts show.

WHY YOU SHOULD NOT NEED “ISOLATE” WITH OSX DEVICES:

With sharing over an OSX port, the VLAN Isolation is complete because the Hypervisors enforce VLAN IDs. Unlike with QDIO, an OSX implementation requires a VLAN ID and a VMAC as assigned by the firmware definitions in the Unified Resource Manager. Remember that OSX is a z/Managed resource and the isolation method that is deployed and required is VLANs.  OSX is a flat Layer 2 broadcast domain using VLANs; technically no Layer 3 routing is required. Deploying ISOLATE can cause unintended networking issues in the Ensemble environment, as you can examine on subsequent charts.

A note about SHARED OSX ports: If you are sharing an OSX OSA port among LPARs and wish to exclude certain devices on the OSX CHPID from being shared, take advantage of the NOTPART keyword in the IOCDS. An LP cannot access a device if the LP is not specified in the device candidate list for the device, even if the LP can access a channel path assigned to the device. The PART or PARTITION keyword specifies the LPs that are in the device candidate list. The NOTPART keyword specifies the LPs that are not in the device candidate list. For example, if a CSS has three LPs (LP1, LP2, and LP3) and you specify NOTPART=(LP2), LP1 and LP3 can access the device, but LP2 cannot. This capability may provide assurances that certain Operating System and TCP/IP stack images cannot define access to the OSX OSA over certain addresses despite the SHARED CHPID. However, if this level of assurance does not satisfy an auditor, you may still exploit the z/Vm or z/OS ISOLATE keyword to enhance the segregation of the Operating Systems sharing the OSX OSA. ISOLATE by itself may satisfy a security auditor or you may use ISOLATE in the interface definition together with the NOTPART definition in the IOCDS. However, improperly deploying ISOLATE can cause unintended networking issues.
Secure Segregation of Traffic over Shared OSD: ISOLATE & Layer 3 Routing

MORE BACKGROUND INFORMATION on ISOLATE with OSD:

Some environments require strict controls for routing data traffic between servers or nodes. In certain cases, the LPAR-to-LPAR capability of a shared OSA port can prevent such controls from being enforced. For example, you may need to ensure that traffic flowing through the OSA adapter does not bypass firewalls or intrusion detection systems implemented in the network.

Dynamic routing protocol implementations with RIP or OSPF require careful planning on LANs where OSA-Express connection isolation is in effect; the dynamic routing protocol learns of the existence of the direct path but is unaware of the isolated configuration, which renders the direct path across the OSA port to the registered target unusable. If the direct path that is operating as ISOLATEd is selected, you will experience routing failures.

Within the IEDN we discourage the use of dynamic routing protocols. For example, OMPROUTE is not aware of the IEDN-to-IEDN forwarding restriction that z/OS enforces and could therefore calculate unusable routes. The IEDN-IEDN forwarding restriction could make its way into other operating systems in the Ensemble; therefore, the use of dynamic protocols should be discouraged.

If the visibility of such errors is undesirable, you can take other measures to avoid the failure messages. If you are simply attempting to bypass the direct route in favor of another, indirect route, you can accomplish this as well with some thoughtful design.

For example, you might purposely bypass the direct path by using Policy Based Routing (PBR) or by coding static routes that supersede the routes learned by the dynamic routing protocol. You might adjust the weights of connections to favor alternate interfaces over the interfaces that have been coded with ISOLATE.

In the visual above we have showed you the static routes that could override the direct routes that OMPROUTE would dynamically learn about.
Effect of Introducing ISOLATE into the Ensemble: 2 Examples

NOTE: ISOLATE is enforced on the INMN network; it is optional on the IEDN network. In z/OS the ISOLATE parameter is coded on the INTERFACE definition; in z/VM ISOLATE is coded on the VSwitch.

In the first visual above you see that the z196 has two LPARs, with each LPAR dedicated to a different company’s production TCP/IP stack: ACORPTCP and BCORPTCP. We are managing the ensemble and keeping the processing secure by ensuring that the traffic from ACORPTCP flows over VLAN10 in the IEDN, and that the traffic from BCORPTCP flows over VLAN11 in the IEDN. ACORPTCP needs to communicate only with VS100 and VS101 in the zBX; BCORPTCP needs to communicate only with VS102 in the zBX. The VLAN enforcement required for Ensemble Networking ensures that the two TCP/IP stacks on the z196 cannot communicate over the IEDN with each other.

In the second visual we show you a different design: ACORPTCP and BCORPTCP are both allowed to send traffic over VLAN10. They are both using the same Virtual Servers in VLAN10 in the zBX, but the TCP stacks themselves are not allowed to communicate directly with each other. For this reason the system administrators have introduced Connection Isolation into the picture by coding ISOLATE on at least one (and possibly both) of the OSX Interface definitions of the two z/OS TCP/IP stacks. Therefore, while permitting communication freely over VLAN10 to Virtual Servers in the zBX, they are not permitting communication over the shared OSX port between the two TCP/IP stacks in the LPARs. This is an unusual network design and one would have to have very good reasons for introducing it. You could eventually see a negative consequence of this design, for example, if you needed to assign a second TCP/IP stack to each company. In such a case, the two TCP/IP stacks of a single company would not be able to communicate with each other at all over the IEDN because of the ISOLATE coding that had already been introduced in the first TCP/IP stack.

WHY YOU SHOULD NOT NEED ISOLATE WITH OSX DEVICES:

With sharing over an OSX port, the VLAN Isolation is complete because the Hypervisors enforce VLAN IDs. Unlike with QDIO, an OSX implementation requires a VLAN ID and a VMAC as assigned by the firmware definitions in the Unified Resource Manager. Remember that OSX is a z/Managed resource and the isolation method that is deployed and required is VLANs. OSX is a flat Layer 2 broadcast domain using VLANs; technically no Layer 3 routing is required. Deploying ISOLATE can cause two stacks sharing an OSX port and using the same VLAN not to be able to communicate with each other over the internally shared OSA path: if the two other stacks need to communicate with each other despite the ISOLATED internal path, then you must deploy a router or a switch solution with IP Firewall Filtering rules to maintain the security and integrity of the Ensemble design.

A note about SHARED OSX ports: If you are sharing an OSX OSA port among LPARs and wish to exclude certain devices on the OSX CHPID from being shared, take advantage of the NOTPART keyword in the IOCDS. An LP cannot access a device if the LP is not specified in the device candidate list for the device, even if the LP can access a channel path assigned to the device. The PART or PARTITION keyword specifies the LPs that are in the device candidate list. The NOTPART keyword specifies the LPs that are not in the device candidate list. For example, if a CSS has three LPs (LP1, LP2, and LP3) and you specify NOTPART(LP1,LP2), LP1 and LP3 can access the device, but LP2 cannot. This capability may provide assurances that certain Operating System and TCP/IP stack images cannot define access to the OSX OSA over certain addresses despite the SHARED CHPID. However, if this level of assurance does not satisfy an auditor, you may still exploit the z/VM or z/OS ISOLATE keyword to enhance the segregation of the Operating Systems sharing the OSX OSA. ISOLATE by itself may satisfy a security auditor or you may use ISOLATE in the interface definition together with the NOTPART definition in the IOCDS. However, improperly deploying ISOLATE can cause unintended networking issues as you may explore further on the subsequent charts.
Where You Might Question the Use of ISOLATE: An Example

**Did You Really Intend to ISOLATE the two BCORP Stacks on VLAN 11 from Each Other?**

1 = Four TCP images on LPARs in the Ensemble are connected to a shared OSA port (OSX CHPID). The two VLAN pairs (VLAN10 and VLAN11) are already isolated from each other because of the VLAN enforcement in the Ensemble network.

-- Technically there is no need to use ISOLATE to separate the two VLANs from each other.

2 = The two ACORP images on VLAN10 can communicate with each other and with Virtual Servers in the zBX that belong to the appropriate VLAN10.

3 = The two BCORP images on VLAN11 can communicate only with Virtual Servers in the zBX that belong to the appropriate VLAN11.

4 = WARNING: The two BCORP images cannot communicate with each other over the shared OSX port because one or both of them have coded ISOLATE on the OSX Interface.

**QUESTION:**

- Did you really intend to keep the two BCORP stacks from communicating with each other over the IEDN?
  - If not, then don’t use ISOLATE or provide an alternate path through a router or use dedicated OSA ports on each. (See next page.)
  - If so, then ISOLATE is appropriate. Question: Why did you not just introduce a separate VLAN ID?

The visual depicts a scenario in which ISOLATE has been coded on both the BCORP LPARs. Perhaps the original intent of coding ISOLATE was to ensure that the ACORP LPARs did not communicate at all across the shared OSX port with the BCORP LPARs. But now look at the consequences: now BCORP111 and BCORP112 cannot communicate with each other at all unless …

-- traffic is forced across a switch and a router, or

-- dedicated OSA ports are introduced as depicted on the next page.

You still need to ask yourself a question, however: If you did not want to allow the two BCORP LPARs to communicate with each other, why did you not just assign them to different VLANs?
Communicate in the IEDN over Dedicated OSX OSA Ports

**Did You Really Intend to Eliminate a Shared OSA Port for the Two BCORP Stacks?**

If four TCP images on LPARs in the Ensemble are connected to a shared OSA port (OSX CHPID) as on the previous diagram, the two VLAN pairs (VLAN10 and VLAN11) are already isolated from each other because of the VLAN enforcement in the Ensemble network.

-- Technically there is no need to use ISOLATE to separate the two VLANs from each other.

2 = The two ACORP images on VLAN10 can communicate with each other and with Virtual Servers in the zBX that belong to the appropriate VLAN10.

3 = The two BCORP images on VLAN11 can communicate with Virtual Servers in the zBX that belong to the appropriate VLAN11.

4 = The two BCORP images communicate with each other over the DEDICATED OSX ports by connecting through the Top of Rack (TOR) switch(es) at Point X:

BCORP111 >>> OSX2 >>> TOR >>> OSX3 >>> BCORP112

**NOTE:** You will need to use a TOR port for every OSX connection.

**Communication Path between BCORP111 and BCORP112:**

- The Communication Path leads from OSX2 to the TOR and then to the switch port that connects to OSX3.

In this diagram the two BCORP virtual servers communicate with each other only through dedicated OSX ports. To support this configuration additional ports on the TOR or TORs are utilized.
More Details GPMP and Management Guest for z/VM
The GPMP agent runs in the Linux guest, like AIX
The data flows thru ZVMLXAPP
ZVMLXAPP is the management guest
It connects to the management network (INMN)
References
References (Basic Manuals, Basic White Papers)

- **zEnterprise Ensemble Planning and Configuring Guide** (GC27-2608-00)
- **zEnterprise System Introduction to Ensembles** (GC27-2609-00)
- **zEnterprise System Hardware Management Console and Support Element Operations Guide for Ensembles** (SC27-2606-00)
- **zEnterprise System Ensemble Performance Management Guide** (GC27-2607-00)

**zEnterprise Network Security White Paper (ZSW03167-USEN-00) and Other Resources**
  - [www.ibm.com/systems/z/resources](http://www.ibm.com/systems/z/resources) (Select “Literature” Entries)

**IBM System z Hardware Management Console Security White Paper**
  - Author Kurt Schroeder ([schroedk@us.ibm.com](mailto:schroedk@us.ibm.com)), Sept. 2008
References (FAQs, Presentations)

- zEnterprise System Frequently Asked Questions (FAQs)
  - www.ibm.com/systems/z/faq (zEnterprise in general)

- IBM zEnterprise System Network Virtualization, Management, and Security (Parts 1 and 2: Overview and Detail)
  - w3.ibm.com/support/techdocs
  - Part 2 also at www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS4160 (Search on keyword “zEnterprise”)
References (Software and Security)

- **z/OS Ensemble Implementation**
  - z/OS Communications Server V1R12 SNA Network Implementation Guide (SC31-8777)
  - z/OS Communications Server V1R12 SNA Network Definition Reference (SC31-8778)
  - z/OS Communications Server V1R12 IP Configuration Guide (SC31-8775)
  - z/OS Communications Server V1R12 IP Configuration Reference (SC31-8776)

- **IPv6 Information**
  - z/OS Communications Server V1R12 IPv6 Network and Application Design Guide (SC31-8885)

- **z/VM Ensemble Implementation**
  - z/VM 6.1 with Small Programming Enhancement (SPE): CP Planning and Configuration (SC24-6083)
  - z/VM Version 6 Release 1 Systems Management Application Programming (SC24-6234-01)

  - [www.redbooks.ibm.com](http://www.redbooks.ibm.com)

- **Security on the IBM Mainframe**, SG24-7803-00 Redbooks®, published 30 April 2010
  - [www.redbooks.ibm.com](http://www.redbooks.ibm.com)

  - [www.redbooks.ibm.com](http://www.redbooks.ibm.com)
IBM® ATS zEnterprise™ System Technical Roadshow

References (Hardware)

- **zBX Publications**
  - zBX Service Guide GC28-6884-01
  - zBX Installation Manual (2458-002) GC27-2610-00
  - zBX IMPP (2458-002) GC27-2611-00
  - zBX Service Education SE245800
  - zBX Safety Inspection (for Model 2) GC28-6889-00
  - IBM License Agreement for Machine Code SC28-6872-00
  - Systems Environmental Notices and User Guide Z125-5823-02
  - Systems Safety Notices G229-9054-02
- **Redbooks (www.redbooks.ibm.com)**
  - IBM zEnterprise Technical Introduction, SG24-7832
  - IBM zEnterprise Technical Guide, SG24-7833
  - IBM zEnterprise Configuration Setup, SG24-7834
  - IBM zEnterprise Platform Management, SG24-7835
  - IBM System p® Advanced POWER Virtualization Best Practices, redp4194
  - IBM BladeCenter JS12 and JS22 Implementation Guide, SG24-7655)
- **zBX 2458-002 SAPR Guide**
  - SA10-006
  - 2458 TDA Confirmation Form
- **System z and zEnterprise**
  - Input/Output Configuration Program User's Guide for ICP IOCP, SB10-7037-08
Questions

Thank You!

gdente@us.ibm.com

The Future Runs on System z
Demo of Unified Resource Manager
To manage an ensemble requires a new HMC role. The traditional role of an HMC is to manage one or more System z server. In addition to the traditional functions available on the HMC, the new role includes management function to control an ensemble. Up to eight z196 and eight zBX attached to those servers can be managed as an ensemble by an HMC. A Unified Resource Manager suite installed on an HMC is required to manage an ensemble.

After we have logged into the HMC, we are presented with the WELCOME screen, which is highlighted in the Navigation Bar on the left, and whose contents are displayed in the screen area to the right. We are using the “Tree Style” under this userid; Tree Style instead of the Classic HMC style.

The z196 has been enabled for Ensemble and therefore the Navigation Panel on the left shows the entry for “Ensemble Management.” We see an expanded view of what is available under Ensemble Management: The name of the Ensemble (ATSENS1), the single Node or Member of the Ensemble, named TSYS, and a view of Workloads that have been defined for the Ensemble.

We also see the HMC Management entry in the Navigation Bar and select it next.
The HMC and SE allows you to choose the interface style you prefer:

**Tree style user interface:**

– Hierarchical views of system resources and tasks using drill-down and launch-in-context techniques to enable direct access to hardware resources and task management capabilities.

**Classic style user interface:**

– The original user interface. It has an object-oriented design. Through this design, you can directly manipulate the objects (such as CPCs or images) that are defined to the HMC. Be aware of any changes to the hardware status as they are detected.

If you are performing HMC Management, you can Configure the User Settings to use the Classic style or the Tree Style. Look at the next screen that shows this selection panel.
The User Interface Style

Choose between:

- Classic Style
- Tree Style

Cancel out of this screen to return to the previous screen. From that screen select the “What’s New” item.
Notice that there is an entry for “What’s New (at 0).” This contains the changes in the new HMC management panels. We select New Security Roles at the bottom of the list (at 1).
The new security roles include User IDs, which you can continue customizing, new Task roles and new Resource Roles, all of which you can customize in the name of security and granular, controlled access to the Ensemble resources and actions.

Select FINISH to return you to the initial screen. Then select the Ensemble Management entry from the left-hand navigation bar.
Having selected Ensemble Management from the Navigation Panel, we now see begin to examine the four areas of the screen:

1. The Task Bar: Shows current tasks that are running
2. The Navigation Panel: Shows the high-level Management areas available to us
3. The Work Panel: Shows the resources in the Ensemble
4. The Task Pad: Shows the tasks available to us

We have expanded the view of the entries by clicking on the PLUS sign in the Work Panel. (A) Note at (B) that we can obtain an Ensemble Management Guide that provides us wizards and tutorials for working with the Ensemble and defining its objects. Select the Ensemble Management Guide.
This Guide steps you through the tasks you must complete to create a working Ensemble. Close this window to return.
z/OS LPARs count as one Virtual Server. z/VM LPARs may contain several Virtual Servers, as can zBX Power Blades. After viewing entries here, return to the Ensemble Tab.
Select the Ensemble name and then, in the Task Pad, browse through the screens for Configuration. But end up at “Manage Virtual Networks” (A). See next screen.
Virtual Networks Created in the Ensemble

Select ISAOpt VLAN (161)
Select ACTION "DETAILS"
TOSP11 is on this VLAN.

At "Manage Virtual Networks" (A): Select 161, and then view details of this VLAN from the Action Pulldown. TOSP11 (B) is on the VLAN ID. Cancel and Close out of the displays.
The contents of the box at (C) have been pulled into the center of the slide due to space restrictions. In reality they are part of the Task Pad.

We have selected the node TSYS in the Work Panel. Notice the carets next to the name of TSYS – these expand into a pop up window with the tasks we can execute for TSYS. These are the same tasks we would be able to select from in the Task Pad at 4 or in the Tasks Pull-down at (D).

Let’s look at the details of the TSYS node by highlighting the entry itself or by selecting “CPC Details” from the Task Pad.
When we arrive at the Details Panel we select the Networking and then the zBX Information. Cancel out of these screens.
Select the Configure Top-of-Rack Switch from the Task Pad under Configuration (E). Leave the first TOR highlighted and press OK to view the configuration of this switch.
Review fields on the screen, noting the VLAN Modes on INTERNAL and EXTERNAL connections, the VLAN connections to the external customer network, and the security fields for MAC address filtering.
Return to Ensemble Management and select the Virtual Servers Tab (F). Select the radio button for the TOSP11 Virtual Server (z/OS). Select the task for Operating system messages in any of multiple ways: From the TASKS pulldown in the Work Panel (G); from the carets next to the Virtual Server name of TOSP11 (H); or from the Daily Tasks in the Task Pad (J). From the operating system screen, if you are authorized, you can enter commands. See next panel.
Display commands like:

- D OMVS,PFS to show the IPv6 physical file system
- D NET,VTAMOPTS,OPTION=ENSEMBLE
- D TCPIP,,N,HOME to show the new interfaces created for Ensemble
- D TCPIP,,OSAINFO,INTFNAME=OSX2C4
- D TCPIP,,OSAINFO,INTFNAME=EZ6OSM01

Always press ENTER or SEND to send the command.
To create a new Virtual Member, one path you may use is to start with the Configuration selection in the Task Pad Area. Note how we have highlighted ATSENS1 in the navigation panel on the left. From this screen, select “Manage Virtual Networks” in the Task Pad.
From the Action Pulldown, we select “New Virtual Network...”
Fill in the name and description of the new virtual network and then assign the VLAN ID that fits into your network design. Press OK.
VLAN1030: Now part of the Ensemble

Note how at this stage the VLAN is Inactive.
Select “Add Hosts to Virtual Network…” and Unified Resource Manager begins collecting a list of available Virtual Servers to add to this VLAN as appropriate.
Select the necessary HOST (OSX or IEDN Attachments) and Press OK.
Top-of-Rack (TOR) Port Numbering
BLADE CENTER 2458-002 IEDN TOR FRONT SWITCH CONNECTIONS

- **SFP+**: 10GbE Optical SR or LR Pluggable Modules.
- **DAC**: 10GbE Direct Attach Cables.

- SWITCH JACK PLUGGING RULES AND SWITCH ZONES:
  - J00 - J07 are SFP+ reserved for Host OS IEDN connections.
  - J08 - J22 are DAC reserved for BC IEDN, SM07/SM09 connections.
  - J23 - J25 are 1 Meter DAC for Switch to Switch on z196 GA1
  - J26 - J30 are SFP+ reserved for zBX to zBX IEDN connections.
  - J31 - J39 are SFP+ reserved for customer (PINK) IEDN connections.
  - J40 Console Port
  - J41 IEDN Switch Management Port

- DAC LINK CABLES ARE 1, 3, & 7 METERS

© 2011 IBM Corporation
End of Topic