10415:
SHARE: Atlanta 2012

"Connectivity to the zBX and Routing in the zEnterprise Ensemble"

Authors:
Gwen Dente, IBM Advanced Technical Support, Gaithersburg, MD (USA)
Friedrich Michael Welter, IBM STG Systems Software Development, Boeblingen, Germany

Thursday, March 15, 2012: 1:30 PM-2:30 PM
Chestnut (Omni Hotel CNN Center)

Abstract

- Connecting the External Customer Network to the Ensemble by attaching to ports in the zBX Top-of-Rack (TOR) switch is simple to do, but it is not the same as connecting any external device to just any Layer 2 switch. When connecting to the Intraensemble data network through the TORs, **the IEDN VLANs must terminate at a ROUTED endpoint on the external platform adjacent to the zBX**. (A ROUTED endpoint is also known as a "Layer 3" endpoint.) **The biggest mistake you can make is to attempt a SWITCHED connection – that is, one that relies on Layer 2 switching protocols.**
- Such attempts are likely to fail because the zBX TOR has been configured to expect Layer 3 routed connectivity and is incompatible with typical Layer 2 switching protocols.
- **Related SHARE presentations in Atlanta:**
  - **10823:** zEnterprise System - Network Architecture & Virtualization Overview - Part 1 of 3
  - **10824:** zEnterprise System - z/OS IEDN Network Design & Implementation - Part 2 of 3
  - **10724:** zEnterprise System - Secure Networking with the zEnterprise Ensemble - Part 3 of 3
- For further information about networking with the zEnterprise System Ensemble, consult the extensive information in **IBM® zEnterprise™ System Network Virtualization, Management, and Security (Parts 1 and 2 - Overview and Details)** at:
  - **http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS4160**
- **For other assistance,**
  - **Please work with your IBM representative, who may open a TechXpress request to consult with the IBM Advanced Technical Support zEnterprise Communications Server Networking Team:**
    - http://techsales4.austin.ibm.com/tsna/techxpress.nsf/request.html

Acknowledgments:

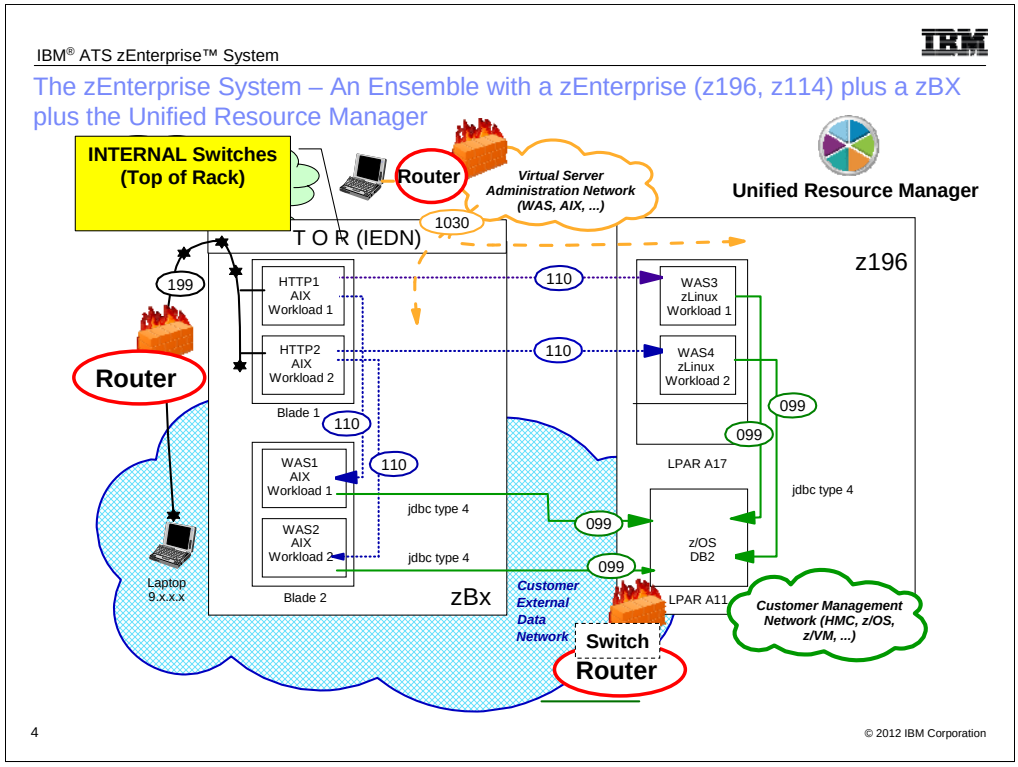
- **Faton Avdiu, IBM Systems &Technology Group, Client Advocacy & System Assurance Network Engineer - (Poughkeepsie)**
- **Patty Driever, IBM System z Hardware Brand Technology Development (Poughkeepsie)**
- **Jerry Stevens, IBM Enterprise Networking Solutions Architecture, Strategy, and Design (Raleigh)**

Agenda

- Review of Ensemble Environment
- Routing protocols in the Ensemble
- Configuring Connectivity to the zBX from the External Customer Network
- * Routed Connectivity (aka “Layer 3 Connectivity”) - Recommended Configurations
- ** Routed Connectivity with Virtual Interfaces
- Appendix: Access vs. Trunk Mode and Configuring the Top-of-Rack Switches

*** Recommended configurations have been tested in IBM labs. (See Page 28.)**

**** Available configuration options should be validated by the customer to verify that certain vendor-specific protocols and implementations abide by the rules for connectivity to the zBX from the external customer network. (See Page 34.)**



When a zBX is part of the zEnterprise Ensemble, the Top of Rack switches (TORs) replace any externally attached switch. That is, we use INTERNAL switches to reach the virtual servers within the Ensemble via the zBX and not EXTERNAL Switches

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 L2 switch. (Certain configurations with L2/L3 switches can be tolerated. See further descriptions in this document.)

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;

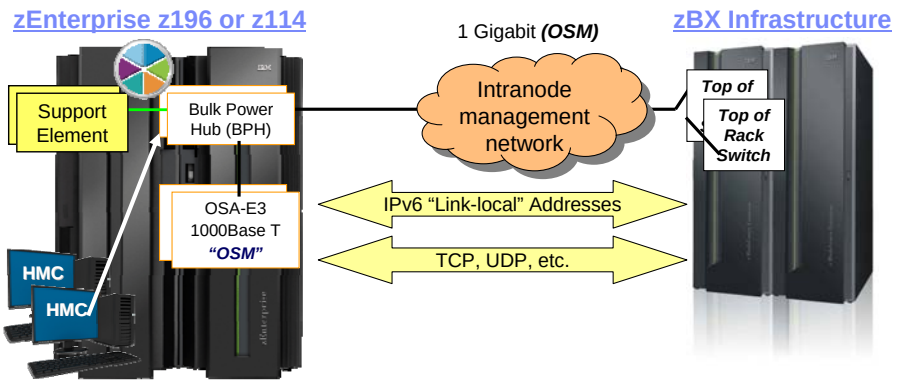
3) The simplified configuration and security enforcement points of the Top-of-Rack switches that extend into the external network.

IMPORTANT Reminder: The TOR is not just any switch .. the switch ensures that possibilities for LAN collisions and for mis-configuration do not impinge on the security of the network because certain standard switch functions like the exchange of Layer 2 messages have been disabled. – This is another reason why the customer cannot replace the existing TOR switch for a different one; Unified Resource Manger integrates with the IBM-provided 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, follow best practices, 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.

Routing in the Ensemble

IBM® ATS zEnterprise™ System

The Intranode Management Network: A Closed, Secure Network



- **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.

6

For reliability in an Ensemble, you must order redundant switches and redundant OSA-E3 and OSA-E4S adapter 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/z114 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.

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 stack 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. Z/VM with Virtual Machines that are Virtual Servers MUST be connected to the INMN. (The z/VM 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.

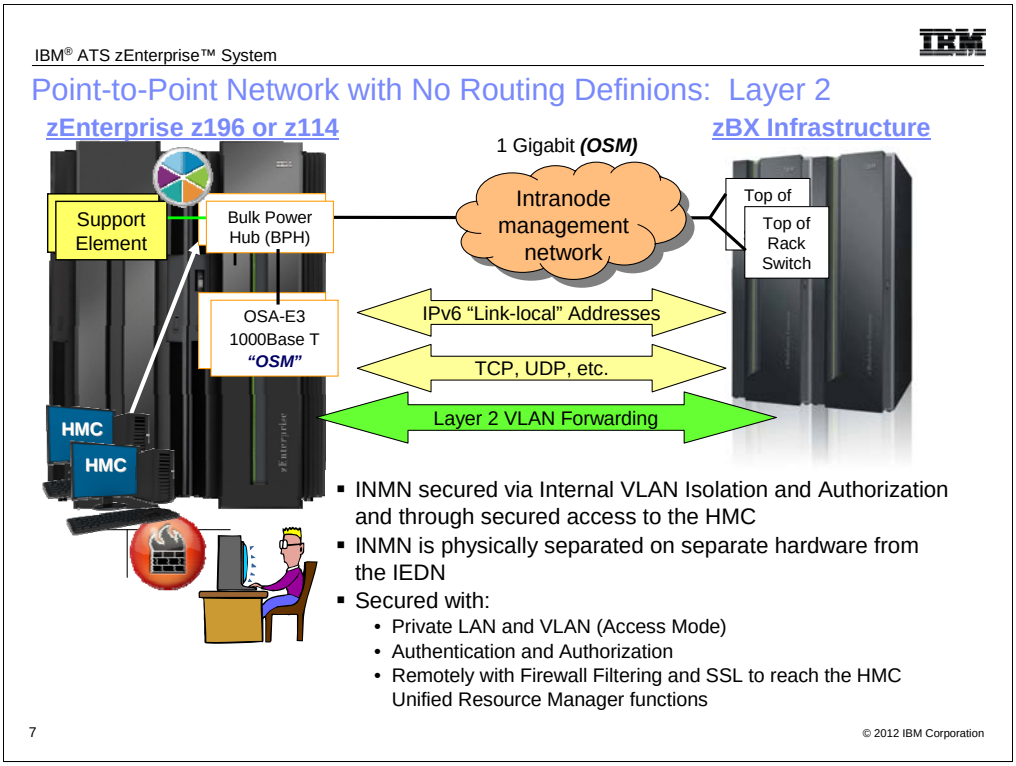
Cabling Specifications:

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/z114 LPARs cannot communicate with each other over the shared OSM ports.

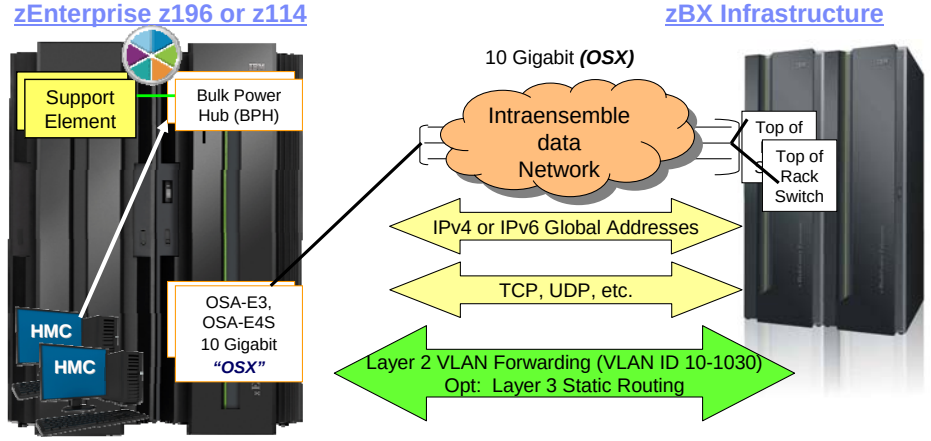
Customer Requirement: To access the INMN through a Network Management Application under secured conditions.

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.

IBM® ATS zEnterprise™ System

The Intraensemble Data Network: Closed or Open Network



- **Intraensemble data network (IEDN)**
  - 10 Gigabit OSA-Express3 or OSA-Express4S — 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

8

** 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, then they 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 and OSA-E4S adapter 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 or OSA-Express4S--- 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)

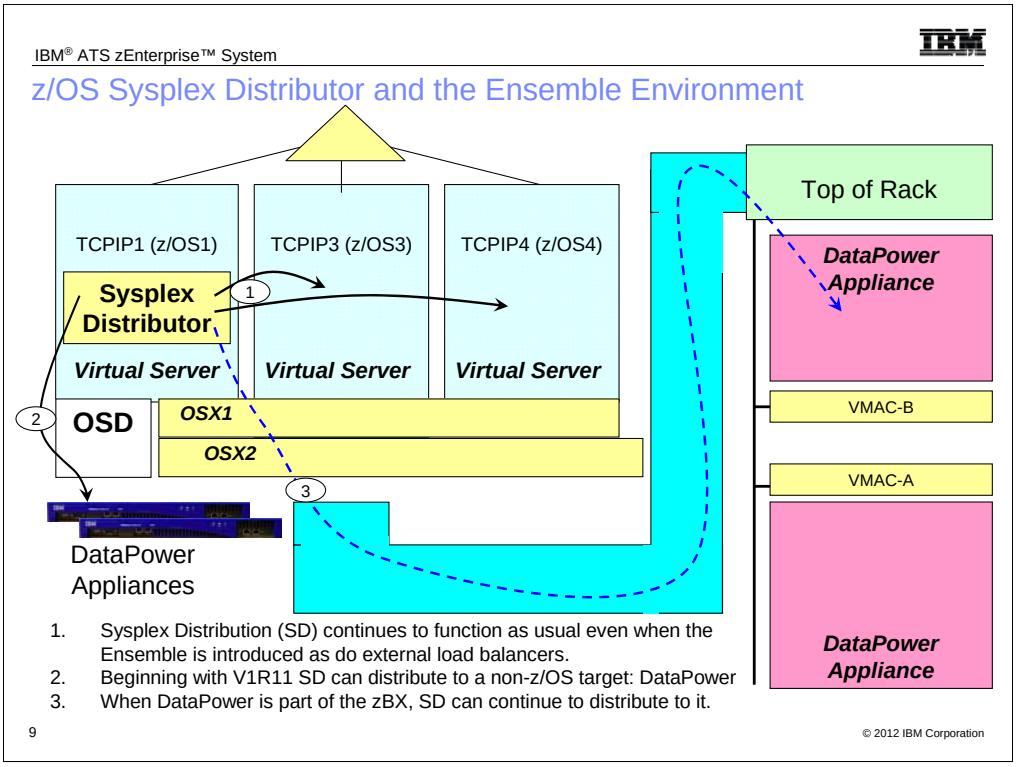
There are eight OSA adapters (16 OSA ports) needed for maximum configuration in a node. Only the first pair of OSA cables is required to be connected to the managing zEnterprise.

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.

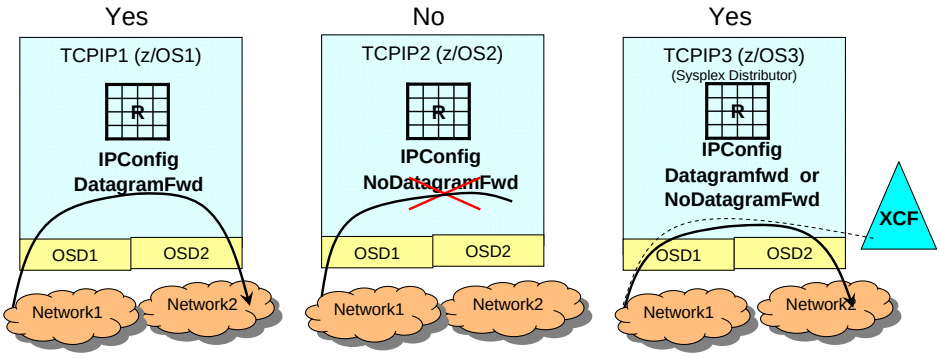
Can implement with static routing. (Introduce dynamic routing only with a carefully designed dynamic routing environment. Due to security implementations in the hypervisors you could find yourself with invalid dynamic tables for OSPF.)



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. (Note: The table in the z/OS Image represents a routing table.) 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.

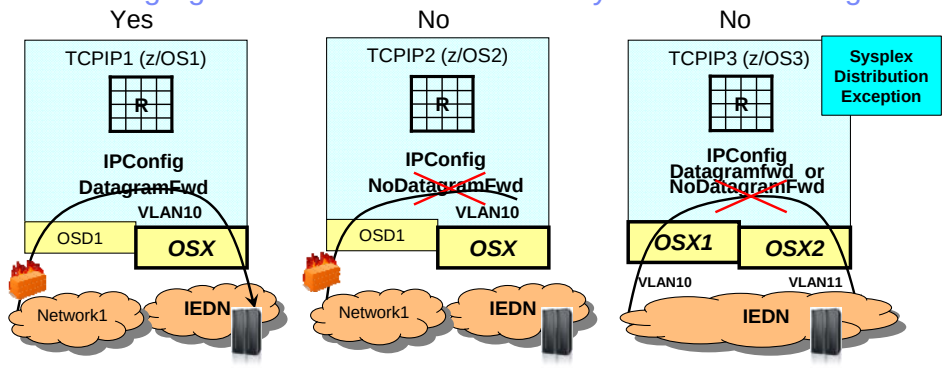
You may disable IP Forwarding in a TCP/IP stack with IPCONFIG NODATAGRAMFWD.

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.

IBM® ATS zEnterprise™ System

Secure Segregation within an IEDN: No Layer 3 IP Forwarding



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.
    1. Forwarding occurs only between Interfaces with identical VLAN IDs.
3. 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.)
4. However, z/OS does not permit Layer 3 routing between OSX or IEDN interfaces at all.
    1. IEDN security has disabled Layer 3 Routing between OSX interfaces. (*Visual 3*)
5. ***Other operating systems (z196/z114 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.”

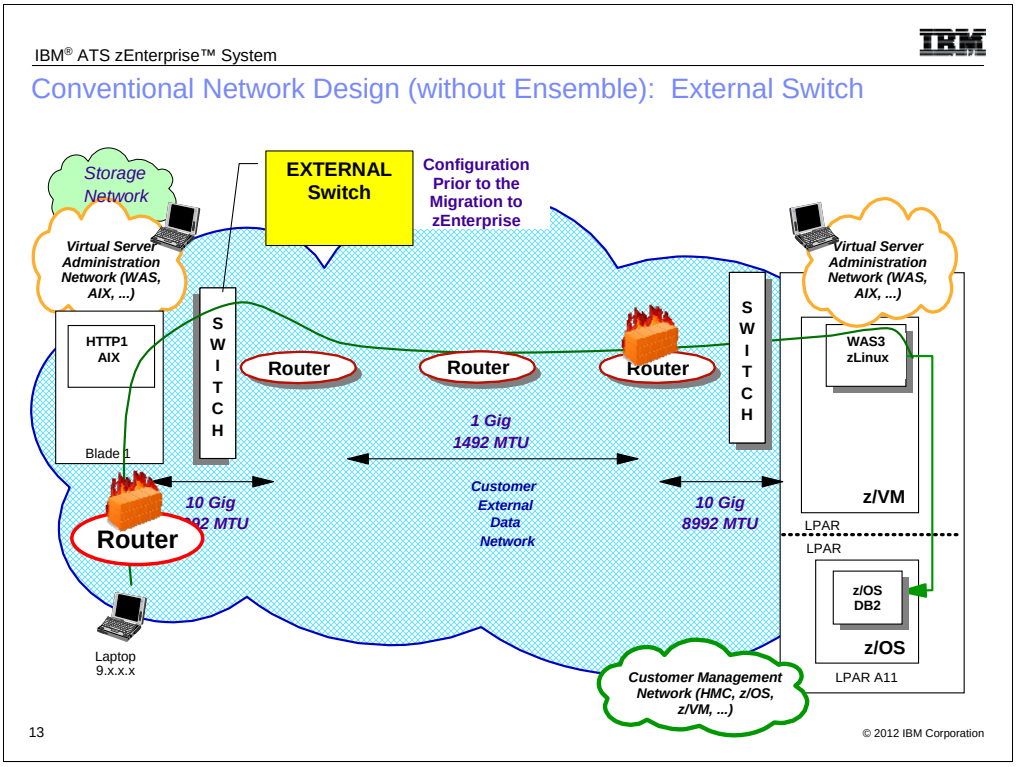
As you see in Visual #1, as long as routed traffic is permitted with IPConfig DatagramFwd and with any potential IP Filtering, then traffic entering an Ensemble Virtual Server by means of a non-IEDN path may be routed over the IEDN OSX OSA port into the IEDN. Visual #2 shows you that IP forwarding is 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.)

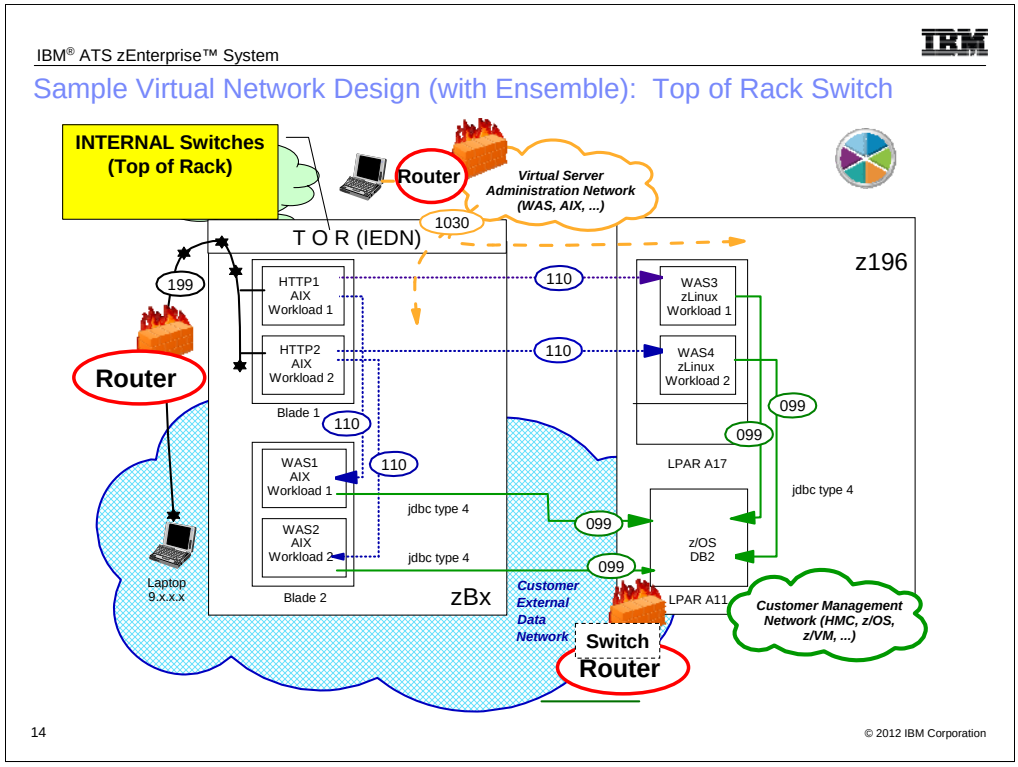
As described IP Forwarding has been disabled in z/OS, but other operating systems (z/VM, for example) in the z196/z114 or in the zBX may indeed still permit IP Forwarding to be able to route between separate VLANs within the IEDN in the same stack.

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.

External Connectivity to the Ensemble



The types of networks you are probably familiar with could look something like this conventional design. The switch attaching to the Blade Center on the left is EXTERNAL to the blade itself. This changes when you migrate to an Ensemble design that includes a zBX.



As you have seen earlier in this presentation, the typical design you saw on the previous page could end up looking like this if it were to be implemented in the Ensemble. There is no EXTERNAL Switch any longer that connects to the zBX – the Switch is now part of the zBX and is represented by two Redundant Top of Rack switches.r

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 L2 switch. (Certain configurations with L2/L3 switches can be tolerated. See further descriptions in this document.)

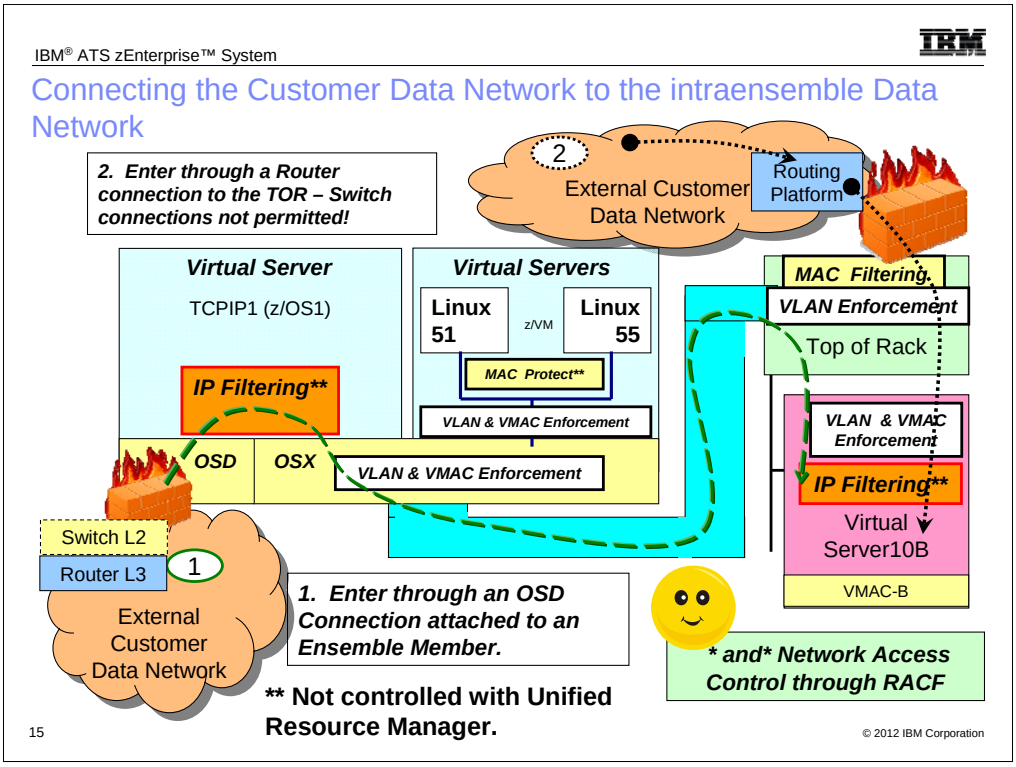
The **differentiators** between the previous visual of a traditional network and this visual with the Ensemble Network depicted in this visual are:

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 mis-configuration do not impinge on the security of the network because certain standard switch functions like the exchange of Layer 2 messages have been disabled. – This is another reason why the customer cannot replace the existing TOR switch for a different one; Unified Resource Manger integrates with the IBM-provided 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, follow best practices, 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.



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. These security measures are outside the control of Unified Resource Manager; note how IP Filtering in z/OS and in the Blade is marked with an asterisk, as is MAC Protect in z/VM. These additional filtering mechanisms require tight control because they are not subject to zManager influence.

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 OSX 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. The external Layer 3 (or Layer2/Layer3) platform cannot bridge into the external network – it MUST ROUTE into it. This requirement for a Layer 3 termination point (a routing termination point) avoids VLAN ID collisions.

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 to avoid network topology or STP topology changes. Other BPDUs, like those for VLAN registration protocols, are also filtered out so that only the Unified Resource Manager can impact VLAN IDs permitted in the IEDN. This TOR filtering together with an external Firewall is to protect the security of the IEDN network to avoid network topology changes.

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.

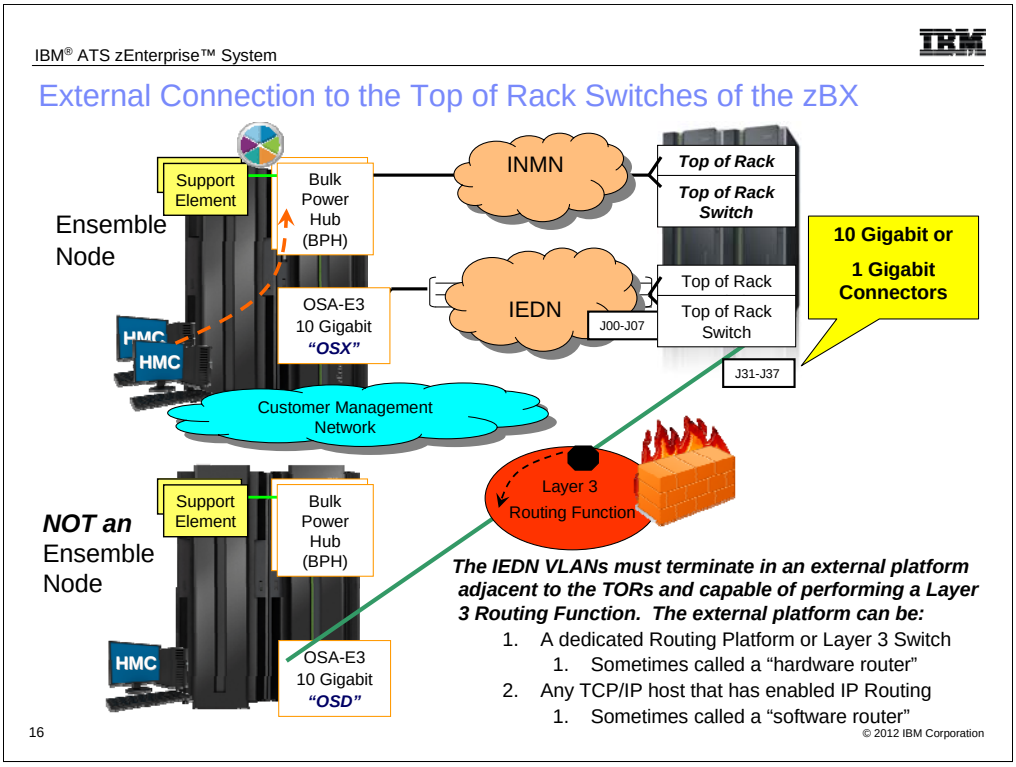
Remember that Security protection is much more than just inserting a firewall along a path. It encompasses all layers of the IP Stack: Application Security Mechanisms (Access Control Lists, Userid and Password checking, mapping mechanisms), Transport Security Mechanisms (SSL/TLS, AT-TLS), IP Layer Security 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 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.

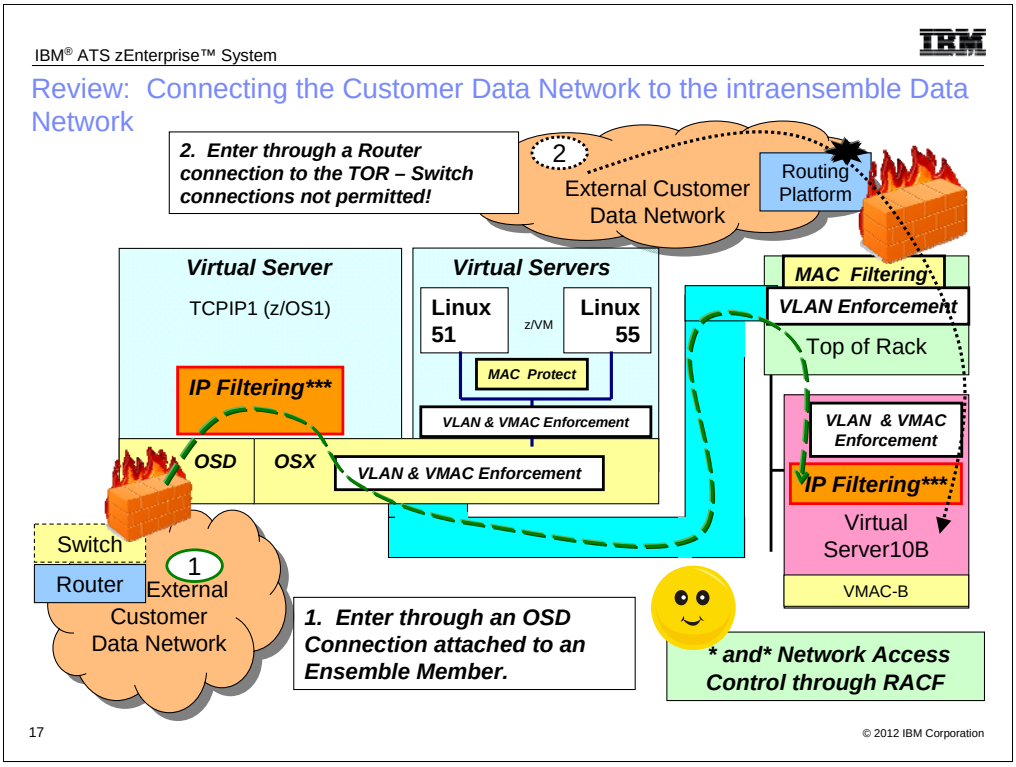
Note on use of VMACs: When an Operating System on z is using layer 2, it performs an ARP and builds Ethernet headers with VMAC. z/OS does not use layer 2 and so ARP is handled by the OSA and the OSA builds the Ethernet header On a VSwitch under z/VM, you might have a Linux guest using Layer 2. In this 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 (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)
  - IP Filtering in Virtual Servers residing on the zBX (may or may not be stateful)
  - Other IP Filtering mechanisms (could be stateful or not)
  - 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.)
  - Firewall in front of TOR in External network (External firewalls are usually "appliance-based" firewalls that – unlike "host-based firewalls" -- are stateful.)
- SERVAUTH Classes: NETACCESS CONTROLS for IEDN
- MAC ADDRESS FILTERING at the TOR
- MultiLevel Security

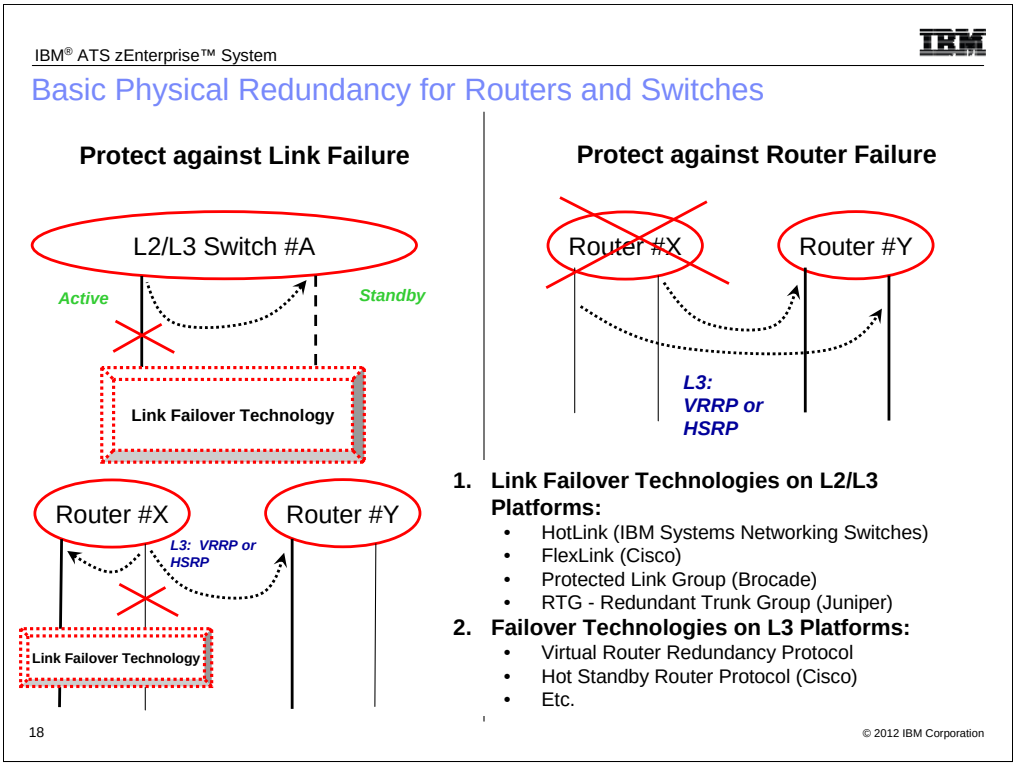


In the visual you see only ONE zEnterprise node. The second System z (either a z196 or another System z platform, which is not a member of the Ensemble) connects over a router (and firewall) from an OSD OSA to the TOR at a port within the port range of J31-J37. Note that the Layer 3 Routing function would most typically be installed in a dedicated routing platform ("router") but could be in any node capable of terminating the IEDN VLANs as a routed endpoint. This is sometimes referred to as connecting to the external platform in Layer 3 mode. In the visual you see that we are implying the existence of a ROUTED termination point in a separate platform adjacent to the TORs. However, this separate platform could have been eliminated and we could have terminated the IEDN VLAN directly in a TCP/IP stack resident on the target Operating System itself. The arrow depicted with the dashed line indicates that we are forwarding data that has arrived at this node using Layer 3 routing mechanisms.



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. **Note that any security implementation that falls outside the strict controls of Unified Resource Manager can be a security weakness if the implementers are not tightly controlled and if there are too many diverse implementers who are not subject to the centralized controls of Unified Resource Manager. Third,** determine how you will secure the external connections. The TOR performs VLAN ID enforcement for connections to servers outside the zBX that are not attached to an OSX OSA port. If appliances are in 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 the IEDN VLANs must terminate at a ROUTED endpoint adjacent to the zBX's TORs. An adjacent switch (layer 2) is not supported. (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.)

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 or IP Filtering 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 zManager/HMC 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 TOR then enforces these VLAN IDs and filtered MACs or VMACs. 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 (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); IP Filtering in Virtual Servers residing on the zBX (may or may not be stateful); Other IP Filtering mechanisms (could be stateful or not); 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.); Firewall in front of TOR in External network (External firewalls are usually "appliance-based" firewalls that – unlike "host-based firewalls" -- are stateful.); SERVAUTH Classes: NETACCESS CONTROLS for IEDN; MAC ADDRESS FILTERING at the TOR; MultiLevel Security



When we want to back up failing components related to a router or a switch, we need to think in terms of 1) protecting against a Physical Link Failure on the platform and 2) protecting against an entire platform failure.

Redundancy Options with One L2/L3 Switch or Router Platform:

Note: This type of failover technology goes by unique names with different vendors and is not standardized.

•2 Interfaces on same VLANs – each attached to a separate TOR in the zBX.

•Interfaces are implemented with a failover technology that does not negatively affect communication with TORs.

•Without a second L2/L3 Switch or without a second Router, this configuration is not optimal since it still includes a single point of failure. Best Practices would dictate that a second L2/L3 or Router platform be added.

Redundancy Options with Two L2/L3 or Router Platforms:

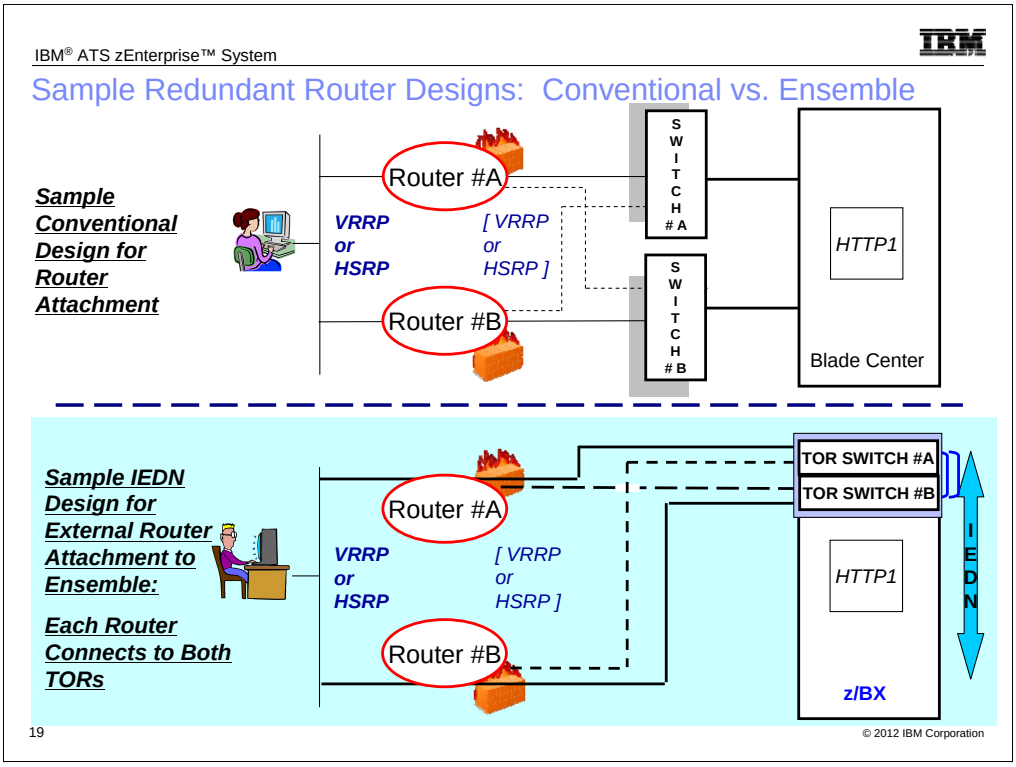
•Single Interface in each platform to the same VLAN and attached to separate TOR in the zBX.

•Interfaces are implemented with a failover technology that does not negatively affect communication with TORs.

•Optional: Each platform can implement link failover within the platform as previously described as long as there is no violation of the restriction about attempting to rely on switch messaging (like STP, any other Bridge Protocol Data Units) for exchanges with the Top of Rack switches (TORs).

If you have a Layer2/Layer3 ("L2/L3") switch and want to build redundancy with 2 links on the same switch (or with one link on each of two interconnected switches), you could use a Layer 2 technology like IBM System Networking Division's (BNT's) HotLinks or Cisco's FlexLinks. Other platform vendors have variations of this same technology as with Hotlink on IBM Systems Networking equipment (formerly BNT) or Protected Link Group on Brocade. In either case, the two links are backed up by placing one in "Active" mode and the second link in "Standby." The value of these settings for attachment to the TOR is that they do not rely on Spanning Tree Protocol, which the TOR would not support. These failover technologies minimize disruption to the network by protecting critical links from loss of data and power. With HotLink, FlexLink, Protected Link Group, or Redundant Trunk Group one port in the group acts as the primary or active link, and the other ports act as secondary or standby link. The active link carries the traffic. If the active link goes down, one of the standby links takes over.

If you are working with a real router – i.e., a purely Layer 3 platform – you can use a Layer 3 failover technology like Virtual Router Redundancy Protocol (VRRP). VRRP can takeover the identity of failed links within a single Layer 3 router or can takeover for the entire failed router by assuming the identity of all the links in the failing platform.



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.

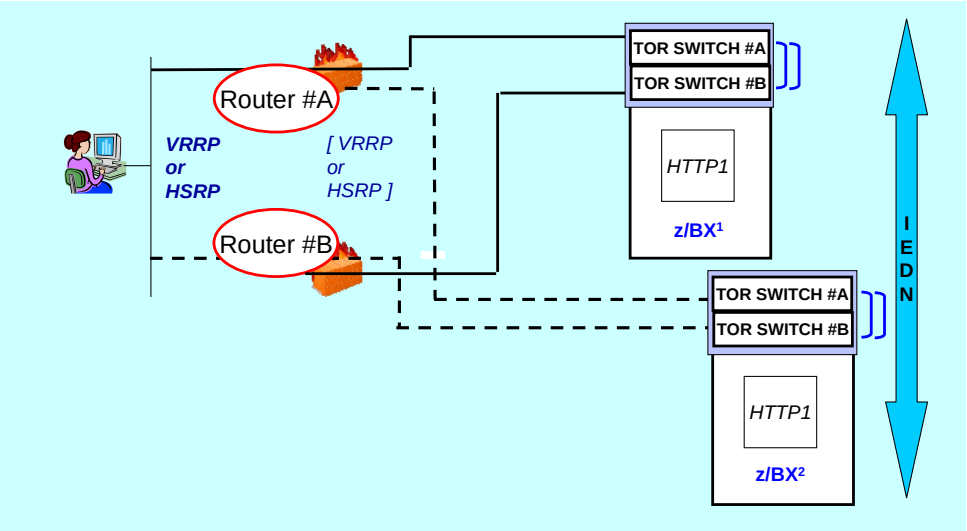
Each Router outside the IEDN may be connected to both TORs, to provide failover to the IEDN regardless of which TOR might fail. This is depicted in the visual at the bottom half of the page. Both routers continue to carry traffic to the IEDN in this fashion.

Alternatively, each router might be connected to a different TOR from the other. If TOR#A fails, Router #A cannot reach the IEDN, but Router #B could take over the traffic load. (The takeover could be triggered by timers of dynamic routing protocols between the router and the external network or through a protocol like VRRP or HSRP.) In either case, failover mechanisms for high availability exist.

Sample Redundant zBX Design for High Availability

Sample IEDN Design for External Router Attachment to Ensemble:

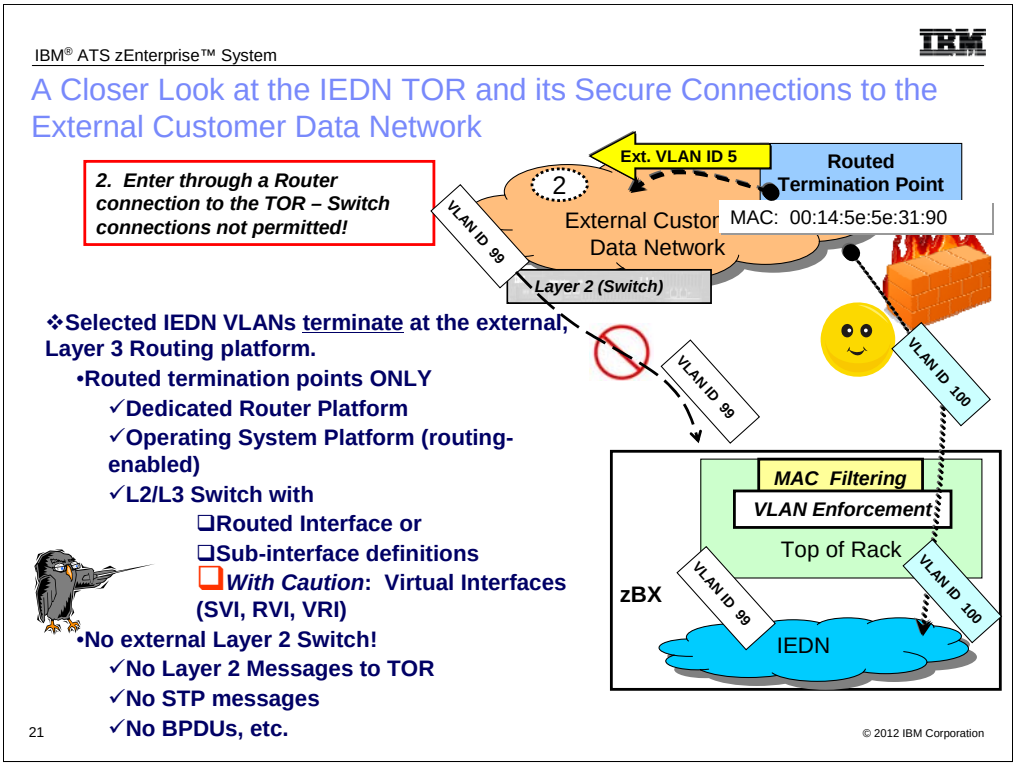
Each Router Connects to TOR A in one zBX and TOR B in a 2nd zBX



This visual shows you how, with multiple zBXs, you might even interconnect routers to two different TORs, but with one TOR residing in one zBX and the other TOR residing in a separate zBX. These two zBXs are part of the same Ensemble and the TORs of zBX^1 are interconnected with the TORs of zBX^2.

In this fashion, even with the loss of a zBX, connectivity with the IEDN is retained.

TOR A and TOR B operate as a single Layer 2 domain. Different configuration options exist to provide the high availability.



Review: This visual emphasizes that VLAN domains in the IEDN may not merge or converge with VLAN domains in the external customer network. The connection (VLAN ID 100) from the IEDN into the external Layer 3 Router must TERMINATE at the Layer 3 platform; the Layer 3 routing function may then route into the customer's VLAN (VLAN ID 5) in his external network. No switch or bridge functions are allowed to interconnect the IEDN with the external network! This safeguard is in place to prevent looping, to prevent movement of a root bridge from one domain's switch to another, and to avoid VLAN collisions. The Layer 3 termination point prevents loops in the network without the need for Spanning Tree Protocol (STP – IEEE 802.1D bridge protocol). STP would require that STP messages (BPDUs) be accepted by the TOR, but the TOR implementation filters out all BPDUs. Therefore, STP is incompatible with a TOR connection to the external network as is link aggregation (802.1ad) into the external network.

About Spanning Tree Protocol: STP (IEEE 802.1D bridge protocol) detects and eliminates logical loops in the network. STP also ensures that the least cost path is taken when multiple paths exist between ports or VLANs. If the selected path fails, STP searches for and then establishes an alternate path to prevent or limit retransmission of data.

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. These are "IP subnet VLANs." 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.

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; this is accomplished because certain standard switch functions -- like the exchange of Layer 2 messages -- have been disabled. This is another reason why the customer cannot replace the existing TOR switch for a different one; Unified Resource Manger integrates with the IBM-provided 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.

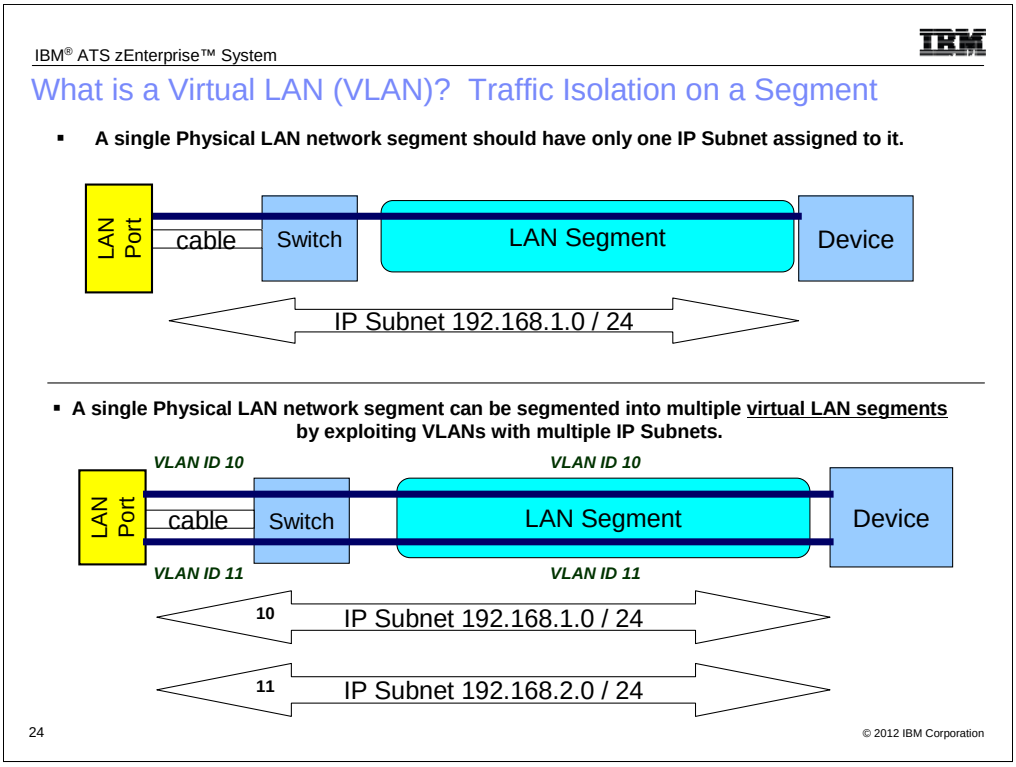
Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

Why We Recommend External Network Routed Termination Point (1)

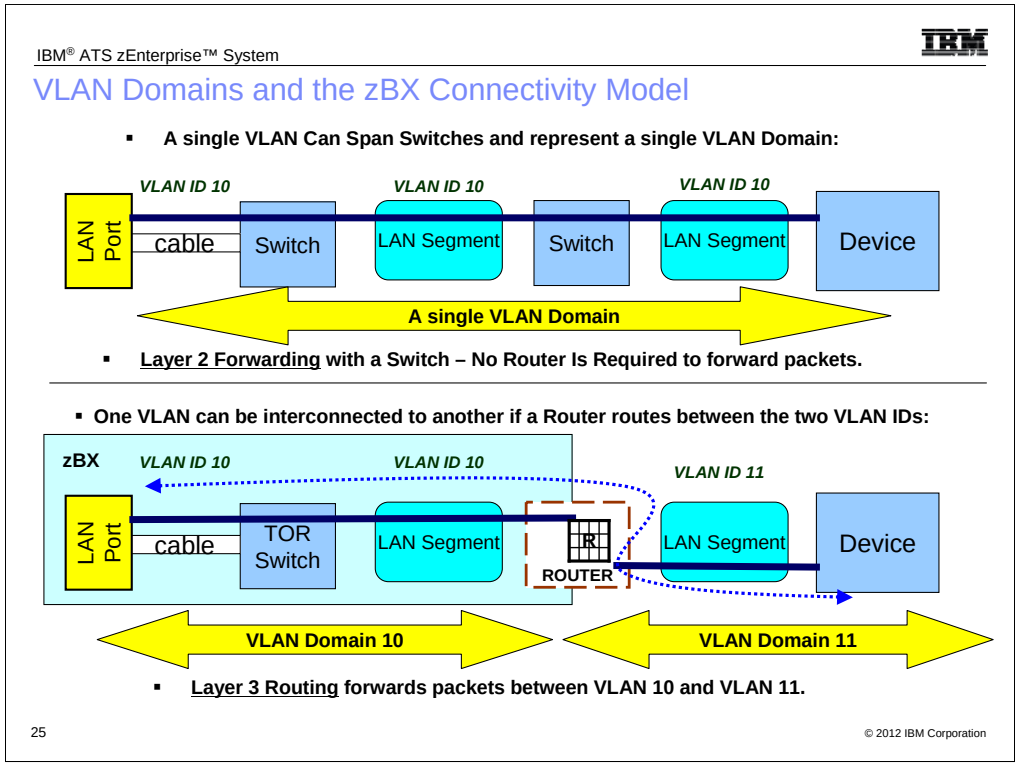
- **Intra-Ensemble Data Network is a 'closed,' flat Layer 2 network**
  - 'Closing' the network is accomplished by **terminating the IEDN VLANs at a ROUTED** endpoint in the external Layer 3 node.
  - Entry into the IEDN is achieved through **Layer 3 IP routing** (either an LPAR or external physical routing platform) into the zBX
  - 'Best practice' from a security and administrative perspective is to place a **firewall/router** prior to entry into the IEDN. This approach provides:
    - Secure isolation/logging/auditing that are typical security requirements when crossing security zones
    - Distinct network administration responsibilities and boundaries (VLANs, VMACs, access controls, etc.)
      - **IEDN VLAN Domains remain segregated from External Customer VLAN Domains.**

Why We Recommend External Network Routed Termination Point (2)

- **In the closed network environment zManager takes total responsibility for:**
  - Network fabric configuration, monitoring, and management
  - Ensuring no virtual MAC or VLAN conflicts (collisions) can occur (and can not be spoofed)
  - Preventing STP packets from passing through the TOR into the IEDN layer 2 LAN segment
  - Identifying, authorizing access, and ensuring all virtual servers within the ensemble can successfully communicate with each other
  - Assuring network high availability is provided (eliminating single points of failures)
  - Single point of Reliability, Availability, Serviceability (RAS) (network diagnostic responsibilities)



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).



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. Connecting separate physical segments of a VLAN by using a Switch creates a single VLAN Domain.

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 traffic that needs to flow across two VLAN domains. A ROUTER operates at what is called “LAYER3” of the IP protocol stack, and examines the contents of the IP Header to determine where to route (forward) the packets across segments of the path. The bottom half of the visual shows you the type of configuration that is required when the IEDN VLAN terminates at the external node adjacent to the TOR switch.

Summary: Avoid the Pitfalls of External Layer 2 Connectivity to the TORs (1)

- **The IEDN VLAN must be terminated at a routed endpoint at the external platform adjacent to the TORs of the zBX.**
  - Isolation through such a routed endpoint provides a clear separation of Layer 2 broadcast domains.
    - Some protocols always provide a ROUTED endpoint:
      - Interface with IP address and optional VLAN ID.
      - Subinterface with IP address and VLAN ID.
    - With proper planning and administration other protocols can provide a ROUTED (Layer 3) endpoint.
      - If incorrectly configured, such protocols present a switched (Layer 2) implementation that merges VLAN domains and violates the requirement for VLAN separation. They are thus not supported for zBX connectivity. **Examples of protocols that must be carefully designed in order to guarantee VLAN isolation are:**
        - **Virtual Interface** implementations like **SVI, RVI, or VRI -- which can provide for either ROUTED or SWITCHED interfaces**
          - (See pages 34-40 for more information on this subject.)

The VLAN termination point outside the zBX must be a Routed (Layer 3) termination point.

By providing layer 3 isolation, there is a clear separation of Layer 2 broadcast domains. Switched Virtual Interface implementations must be carefully planned, since the inappropriate implementation cannot guarantee this separation of Layer 2 (VLAN) broadcast domains.

Any configuration of a connection between an external node and the IEDN TOR cannot rely on Layer 2 messaging protocols at the TOR. Therefore, not supported – among other protocols – are inter-switch protocols whose functions rely on BPDUs. **Examples of such unsupported protocols are:**

Virtual Trunking Protocol (VTP) – its exploitation would limit the allowed IEDN VLANIDs even further

Spanning Tree Protocol (STP) – its exploitation would cause looping

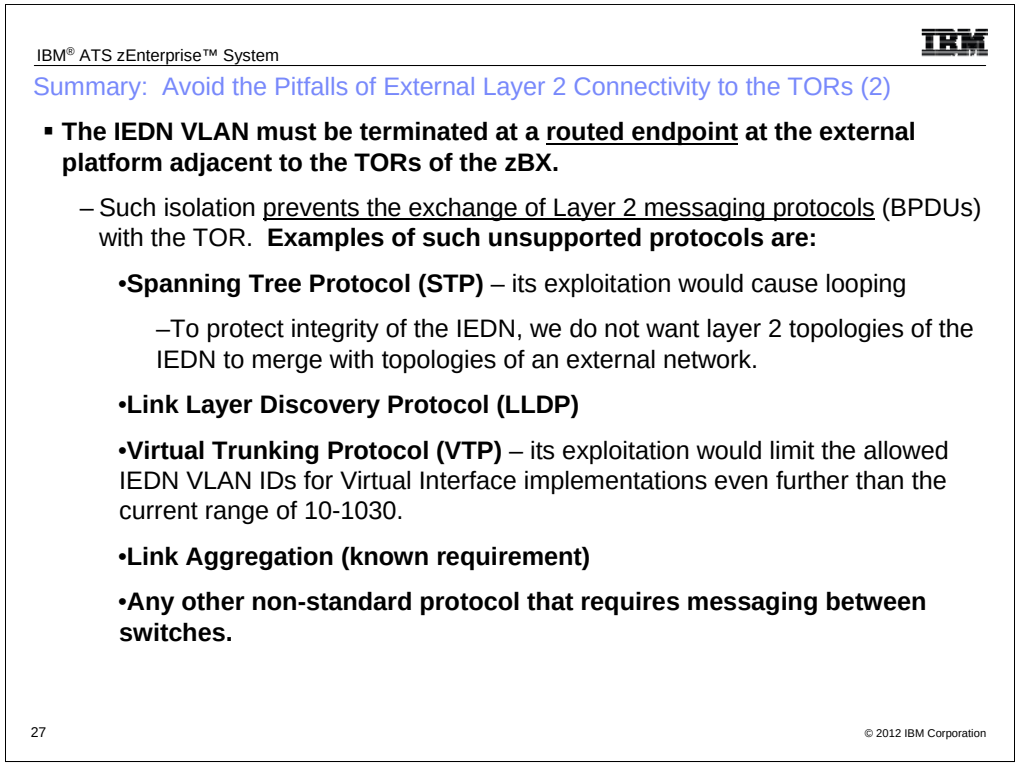
To protect integrity of the IEDN, we do not want layer 2 topologies of the IEDN to merge with topologies of an external network.

Link Layer Discovery Protocol (LLDP)

Link Aggregation (known requirement)

Any other non-standard protocol that requires messaging between switches.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces



The VLAN termination point outside the zBX must be a Routed (Layer 3) termination point.

By providing layer 3 isolation, there is a clear separation of Layer 2 broadcast domains. Switched Virtual Interface implementations must be carefully planned, since the inappropriate implementation cannot guarantee this separation of Layer 2 (VLAN) broadcast domains.

Any configuration of a connection between an external node and the IEDN TOR cannot rely on Layer 2 messaging protocols at the TOR. Therefore, not supported – among other protocols – are inter-switch protocols whose functions rely on BPDUs. **Examples of such unsupported protocols are:**

Virtual Trunking Protocol (VTP) – its exploitation would limit the allowed IEDN VLAN IDs for Virtual Interface implementations even further than the current range of 10-1030.

Spanning Tree Protocol (STP) – its exploitation would cause looping

To protect integrity of the IEDN, we do not want layer 2 topologies of the IEDN to merge with topologies of an external network.

Link Layer Discovery Protocol (LLDP)

Link Aggregation (known requirement)

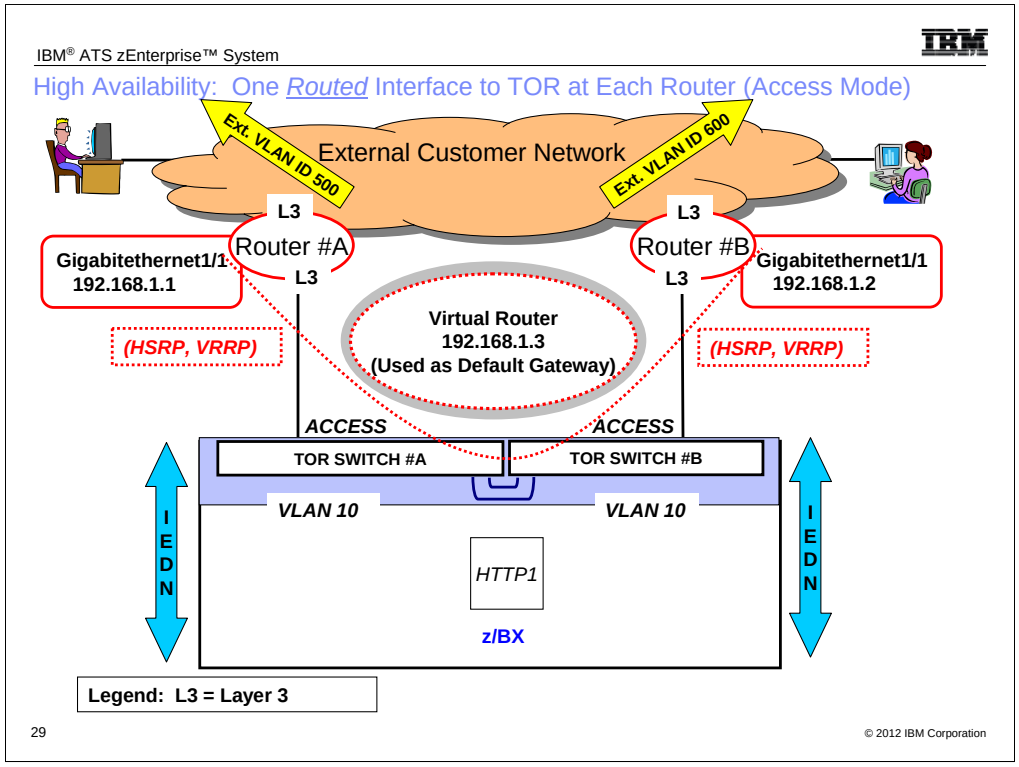
Any other non-standard protocol that requires messaging between switches.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

*Recommended Configurations:
“Routed Connectivity (Layer 3 Connectivity) - Recommended and Supported"

*** Recommended configurations have been tested in IBM labs.**

**** Available configuration options should be validated by the customer to verify that certain vendor-specific protocols and implementations abide by the rules for connectivity to the zBX from the external customer network. (See Page 34.)**



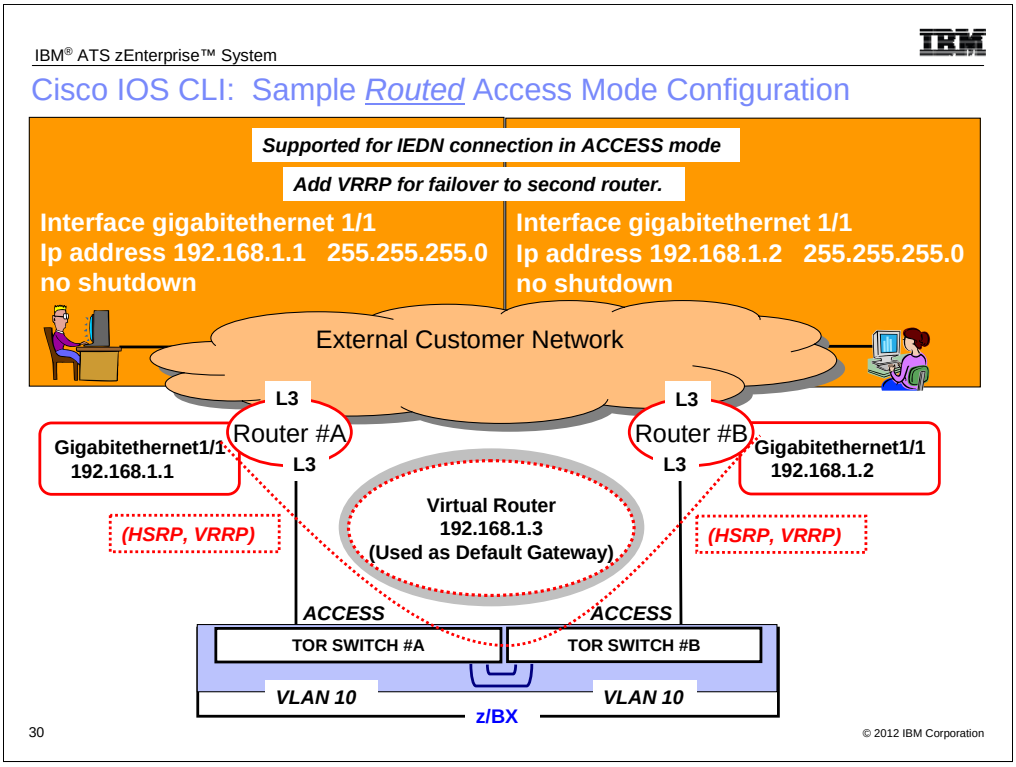
The visual depicts IEDN VLAN ID of 10 which is completely separated from the Customer's External Network VLAN IDs of 500 and 600. This separation of the IEDN domain from the External VLAN Domains is achieved through the termination of the IEDN VLANs at the routed interfaces on the routing platforms.

Two separate Routers

- •One Physical Interface on Each Router
  - •Each Physical Interface is coded with a Routed Interface

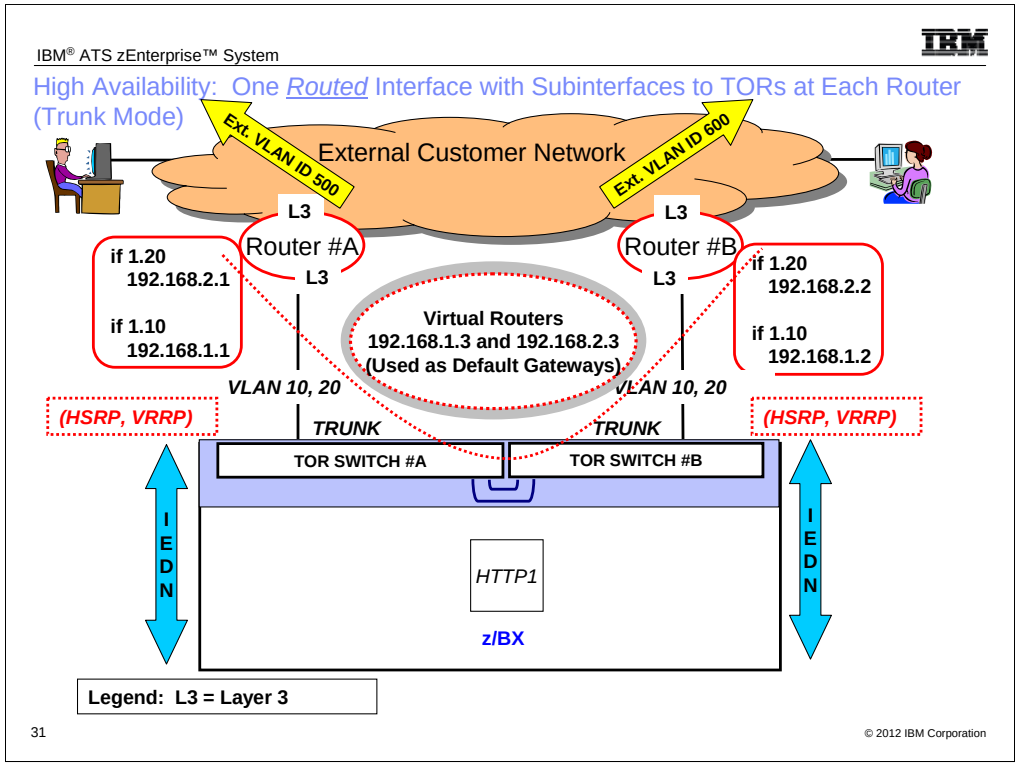
In the event of failure:

Virtual Router Redundancy Protocol (VRRP) or Hot Standby Router Protocol (HSRP) or similar Layer 3 technology backs up a failing interface or failing router. The preferred connectivity for the "heartbeat" messages between VRRP partners is through the TOR path. Although not depicted din this visual, even the failure of the two connections between TOR#A and TOR#B leaves failover paths in place between the two TORs. The failover paths under the control of Rapid Spanning Tree Protocol (RSTP) lead through the High Speed Electronic Switching Modules (ESMs) of the Blade Centers. With a second zBX frame or even second zBX in the scenario, the failover paths between TORs multiply.



http://docwiki.cisco.com/wiki/Cisco_NX-OS/IOS_Interface_Comparison

Recommended: The use of Routed interfaces (either a simple definition or one with Subinterfaces) is the preferred definition for attachment to the zBX TORs. It is impossible to cause a VLAN ID domain merger with this type of configuration.



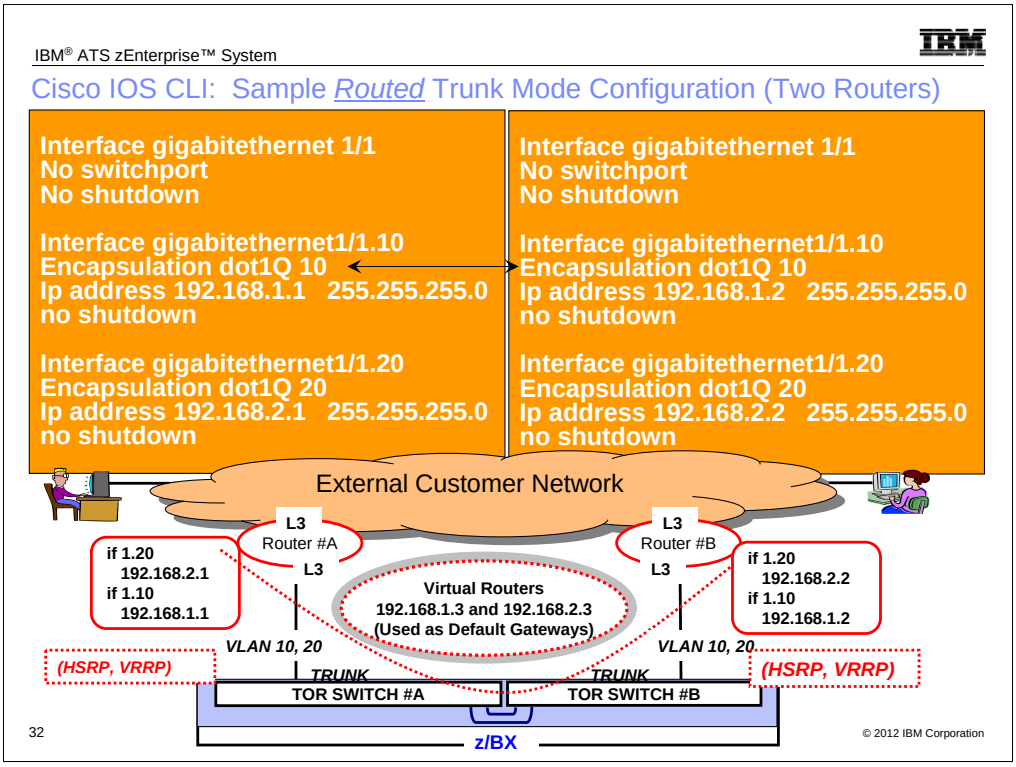
The visual depicts IEDN VLAN IDs of 10 and 20 which are completely separated from the Customer's External Network VLAN IDs of 500 and 600. This separation of the IEDN domain from the External VLAN Domains is achieved through the termination of the IEDN VLANs at the routed interfaces on the routing platforms.

Two separate Routers

- •One Physical Interface on Each Router
  - •Each Physical Interface is coded with two Routed Subinterfaces ***(IMPORTANT: Switched Virtual Interfaces are not supported for connectivity with the TORs.)***
    - •Each Subinterface under a Physical Port represents a different VLAN and a different IP subnet

In the event of failure:

Virtual Router Redundancy Protocol (VRRP) or Hot Standby Router Protocol (HSRP) or similar Layer 3 technology backs up a failing interface or failing router. The preferred connectivity for the "heartbeat" messages between VRRP partners is through the TOR path. Although not depicted in this visual, even the failure of the two connections between TOR#A and TOR#B leaves failover paths in place between the two TORs. The failover paths under the control of Rapid Spanning Tree Protocol (RSTP) lead through the High Speed Electronic Switching Modules (ESMs) of the Blade Centers. With a second zBX frame or even second zBX in the scenario, the failover paths between TORs multiply.

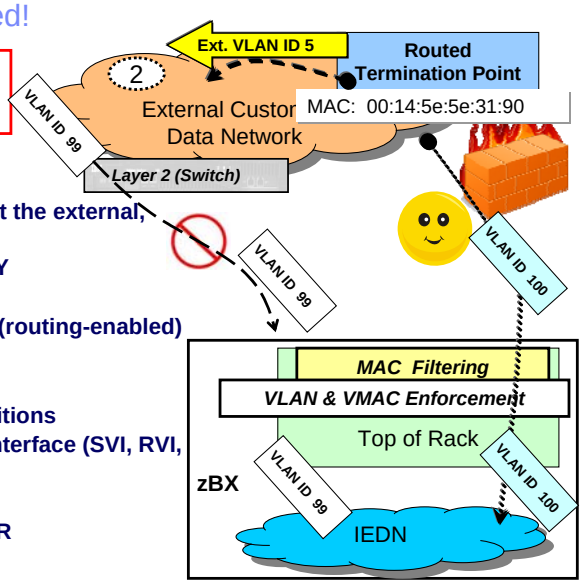


http://docwiki.cisco.com/wiki/Cisco_NX-OS/IOS_Interface_Comparison

Recommended: The use of Routed interfaces (either a simple definition or one with Subinterfaces) is the preferred definition for attachment to the zBX TORs. It is impossible to cause a VLAN ID domain merger with this configuration. The VRRP/HSRP path between the two Routers connects through the TORs. Notice the line "Encapsulation dot1Q …". You must use this line because Cisco defaults to 802.1 ISL protocol, a proprietary protocol which is incompatible with the 802.1q protocol used on the TOR switches.

What is the BIG Mistake You Must Avoid? No Switching Protocol (Layer 2) Messages Permitted!

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



- **Selected IEDN VLANs terminate at the external, Layer 3 Routing platform.**
  - **Routed termination points ONLY**
    - ✓**Dedicated Router Platform**
    - ✓**Operating System Platform (routing-enabled)**
    - ✓**L2/L3 Switch with**
      - ❑**Routed Interface or**
      - ❑**Sub-interface definitions**
      - ❑**Caution: Virtual Interface (SVI, RVI, VRI)**
  - **No external Layer 2 Switch!**
    - ✓**No Layer 2 Messages to TOR**
    - ✓**No STP messages**
    - ✓**No BPDUs of any kind**

Review: You have seen this visual before; we use it again for emphasis of the main point: The IEDN VLAN must terminate at a ROUTED endpoint on the platform adjacent to the zBX. This visual emphasizes that VLAN domains in the IEDN may not merge or converge with VLAN domains in the external customer network. The connection (VLAN ID 100) from the IEDN into the external Layer 3 Router must TERMINATE at the Layer 3 platform; the Layer 3 routing function may then route into the customer's VLAN (VLAN ID 5) in his external network. No switch or bridge functions are allowed to interconnect the IEDN with the external network! This safeguard is in place to prevent looping, to prevent movement of a root bridge from one domain's switch to another, and to avoid VLAN collisions. The Layer 3 termination point prevents loops in the network without the need for Spanning Tree Protocol (STP – IEEE 802.1D bridge protocol). STP would require that STP messages (BPDUs) be accepted by the TOR, but the TOR implementation filters out all BPDUs. Therefore, STP is incompatible with a TOR connection.

About Spanning Tree Protocol: STP (IEEE 802.1D bridge protocol) detects and eliminates logical loops in the network. STP also ensures that the least cost path is taken when multiple paths exist between ports or VLANs. If the selected path fails, STP searches for and then establishes an alternate path to prevent or limit retransmission of data.

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. These are "IP subnet VLANs." 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.

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; this is accomplished because certain standard switch functions -- like the exchange of Layer 2 messages -- have been disabled. This is another reason why the customer cannot replace the existing TOR switch for a different one; Unified Resource Manger integrates with the IBM-provided 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 Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

**** Other Available Options:**
"Routed Connectivity with Virtual Interfaces" -
Supported if Design Adheres to the "Rules":

1. **Routed termination point in external node**
2. **Segregation of IEDN VLAN domains**
3. **No Switching between TORs and external node.**

*** Recommended configurations have been tested in IBM labs. (See Page 28.)**

**** Available configuration options should be validated by the customer to verify that certain vendor-specific protocols and implementations abide by the rules for connectivity to the zBX from the external customer network.**

Correctly Planning Virtual Interface Implementations

Each active switch port on a L2/L3 platform implemented with routed interfaces using SVI, RVI, or VRI must be associated with a unique VLAN ID.

1. **IEDN VLANs can terminate at a Virtual Interface in a L2/L3 platform external to the zBX and its TORs.**
   - **Virtual Interfaces are defined with a Layer 3 IP Address and a VLAN ID**
   - **Virtual Interfaces provide Layer 3 IP connectivity to the TORs and can route between unique VLANs in a L2/L3 platform.**
     - **But … ONLY if the VLANIDs have unique numbers!**
       - **Otherwise traffic will be switched instead of routed.**
2. **Caution: Virtual Interfaces can SWITCH between VLAN segments**
     - **If the IEDN VLAN IDs are identical to the external VLANIDs on the same switch!**
       - **Switched traffic is not supported between an external platform and the TORs.**

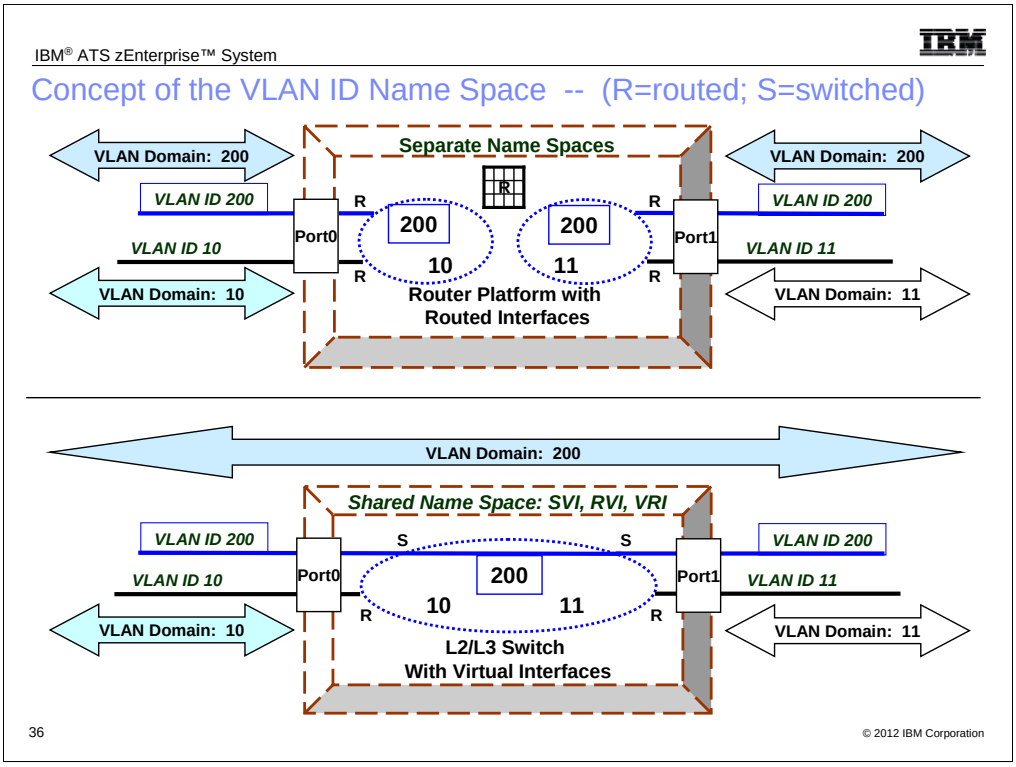
On the previous page you saw that connectivity to the TORs by the external customer network must abide by three rules:

1. **Routed termination point in external node**
2. **Segregation of IEDN VLAN domain**
3. **No Switching between TORs and external node.**

Bullet 1 on this page explains how even an implementation like Switched Virtual Interface, Routed Virtual Interface, or Virtual Routing Interface can abide by all three rules you have learned about connecting to the TOR from the external customer network.

Bullet 2 shows you how an incorrect configuration of SVI, RVI, VRI can lead you to violate those rules and prevent connectivity between the TORs and the external customer network. Always test the Virtual Interface implementation that you plan to use before committing to it in order **to verify that certain vendor-specific protocols and implementations abide by the rules for connectivity to the zBX from the external customer network.**

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces



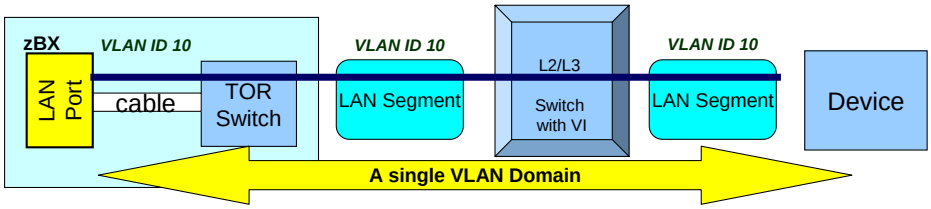
Visual at Top of Page: With Routed Interfaces or Sub-Interfaces defined in a Routing platform or in a L2/L3 switch, separate Name Spaces exist for the VLAN IDs on a per-port basis. As a result we see 4 VLAN Domains. On the left side of the platform we see VLAN Domain 10, VLAN Domain 200. On the right side of the platform we see VLAN Domain 11, and a separate VLAN Domain 200. The domains numbered with 200 are maintained as separate domains even though they have the same VLAN ID number; however, most customers avoid assigning different VLAN domains with the same number although it is possible with separate VLAN ID Name Spaces.

Visual at Bottom of Page: With Virtual Interfaces defined in a Routing platform or in a L2/L3 switch that has implemented SVI, RVI, or VRI, a shared VLAN ID Name Space exists. The logic in Virtual Interface implementations decides through the VLAN ID numbering whether the VLAN endpoints are ROUTED or Switched. If the VLAN IDs are unique on the platform, the Virtual Interface is considered a Routed endpoint. If the VLAN IDs are not unique on the platform and in the shared Name Space, the endpoints are considered SWITCH endpoints. As a result we see 3 VLAN Domains with this Virtual Interface implementation. On the left side of the platform we see VLAN Domain 10. On the right side of the platform we see VLAN Domain 11. Spanning the platform we see a single VLAN Domain #200.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

IBM® ATS zEnterprise™ System

Invalid Virtual Interface (VI) Model (SVI, RVI, VRI)



- ***This Virtual Interface implementation at the external node violates the rules that protect the integrity of the IEDN. The identical VLAN IDs create a single VLAN domain. It is INVALID.***

2. **Caution: Virtual Interfaces can switch (instead of route) between VLAN segments**
   - **If the IEDN VLAN IDs are identical to the external VLANIDs on the same switch!**
     - **Switched traffic is not supported between an external platform and the TORs.**

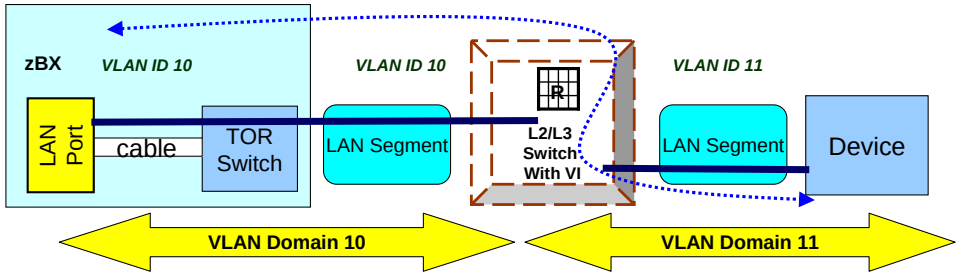
37

IMPORTANT: Always test the Virtual Interface implementation that you plan to use before committing to it since IBM has not been able to test every type of vendor Virtual Interface definition that exists.

The visual at the top of this page represents an INVALID Virtual Interface implementation because the IEDN VLAN ID is identified with the same number as the VLAN ID of the external customer network. With a shared VLAN ID Name Space such a configuration causes the identical VLAN IDs to invoke a switch operation in the VI platform. This configuration is INVALID for connectivity to the TOR.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

Valid Virtual Interface (VI) Model (SVI, RVI, VRI)



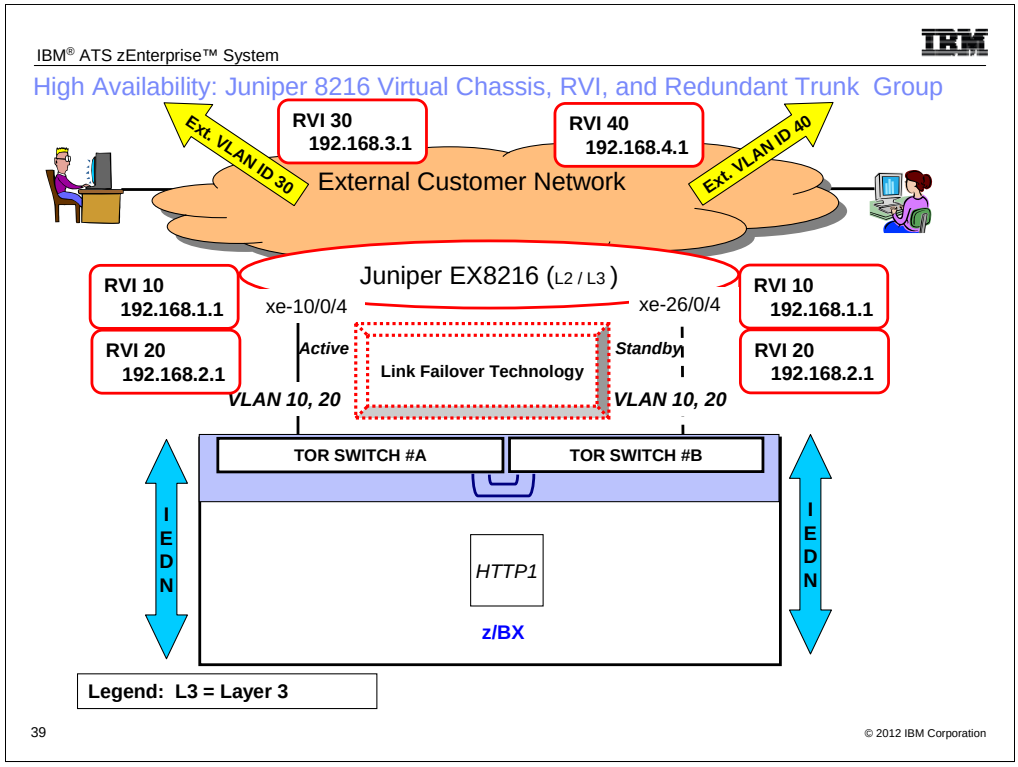
- ***The unique VLAN IDs create two VLAN domains.*** **Layer 3 Routing forwards packets between VLAN 10 and VLAN 11.** ***This is a VALID configuration.***

1. **IEDN VLANs can terminate at a Virtual Interface in a L2/L3 platform external to the zBX and its TORs.**
   - **Virtual Interfaces are defined with Layer 3 IP Addresses and optional VLAN IDs**
   - **Virtual Interfaces provide Layer 3 IP connectivity to the TORs and can route between unique VLANs in a L2/L3 platform.**
     - **But … ONLY if the VLANIDs have unique numbers!**
       - **Otherwise traffic will be switched instead of routed.**

The visual at the top of the page represents what could be a valid configuration for a Virtual Interface implementation on the external node that is adjacent to the TORs of the zBX. Notice how the L2/L3 switch is routing between VLAN IDs that are represented with unique numbers: VLAN ID 10 and VLAN ID 11. The implementation MUST use separate VLAN IDs because Virtual Interface implementations use a shared VLAN ID Name Space.

IMPORTANT: Always test the Virtual Interface implementation that you plan to use before committing to it since IBM has not been able to test every type of vendor Virtual Interface definition that exists.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces



The visual depicts IEDN VLAN IDs of 10 and 20 which are completely separated from the Customer's External Network VLAN IDs of 30 and 40. This separation of the IEDN domain from the External VLAN Domains is achieved through the termination of the IEDN VLANs at the routed interfaces on the routing platforms.

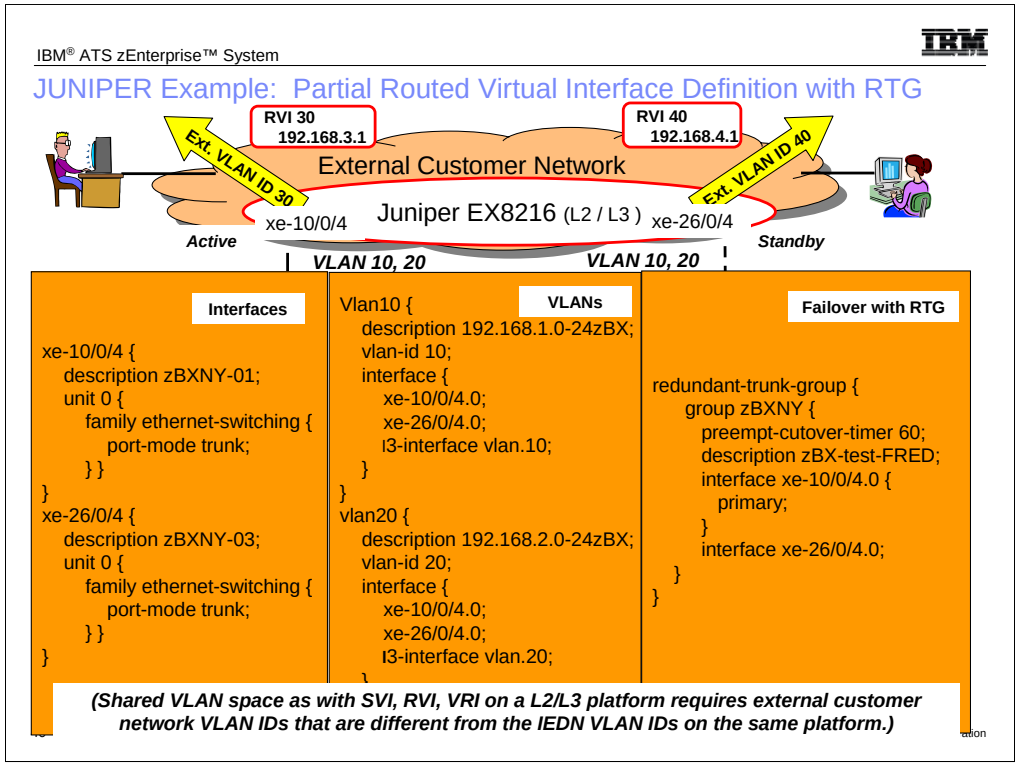
- Two Physical Interfaces on the Routing Platform (Layer 2/Layer 3 switch)
  - Each Physical Interface is coded with two Switched Virtual Interfaces; One physical interface is the Primary and the other is the Standby link. You would need to implement and test a Link Failover Technology to provide high availability and to ensure that the chosen link failover technology does not violate the requirement to avoid the transmission of LAYER 2 Switching Messages to the TOR.
  - The diagram shows that the VLAN IDs in the external customer domain are 30 and 40, which do not conflict with the IEDN VLAN IDs of 10 and 20. Therefore this Virtual Interface implementation maintains the VLAN ID isolation that is required for the security and management design points of the Ensemble.

Switched Virtual Interfaces (also known as Routed Virtual Interfaces – RVI – or Virtual Routed Interfaces – VRI) on Layer 2/Layer 3 Switches -- together with a failover technology like HotLink, FlexLink, Protected Link Group, Redundant Trunk Group – are supported, but **with caution**. The VLAN IDs in the IEDN should be distinct from the VLAN IDs in the external customer network attached to the same switch. Great care must be taken to create an IEDN domain that is separate from any of the layer 2 domains in the external network.This puts an extra burden on the VLAN designer and the administrator of the switch.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

Possible Link Failover Technologies:

- HotLink = IBM Systems Networking (formerly BNT)
- FlexLink = Cisco
- Protected Link Group = Brocade
- Redundant Trunk Group (RTG) = Juniper



Portions of the definition have been elided for purposes of the visual. Virtual Interface implementations (also known as Switched Virtual Interfaces – SVI – or Routed Virtual Interfaces – RVI – or Virtual Routed Interfaces – VRI) on Layer 2/Layer 3 Switches -- together with a failover technology like HotLink, FlexLink, Protected Link Group – are supported, but with caution. **Each active switch port on a L2/L3 platform implemented with routed interfaces using SVI, RVI, or VRI must be associated with a unique VLAN ID. Therefore, with connections of the same VLAN ID into two different TOR ports, you must implement a link failover technology that leaves only one of the two connections active and the other in standby mode. (Otherwise you would have two active ports with the same VLAN ID in the attached platform and cause a loop.)**

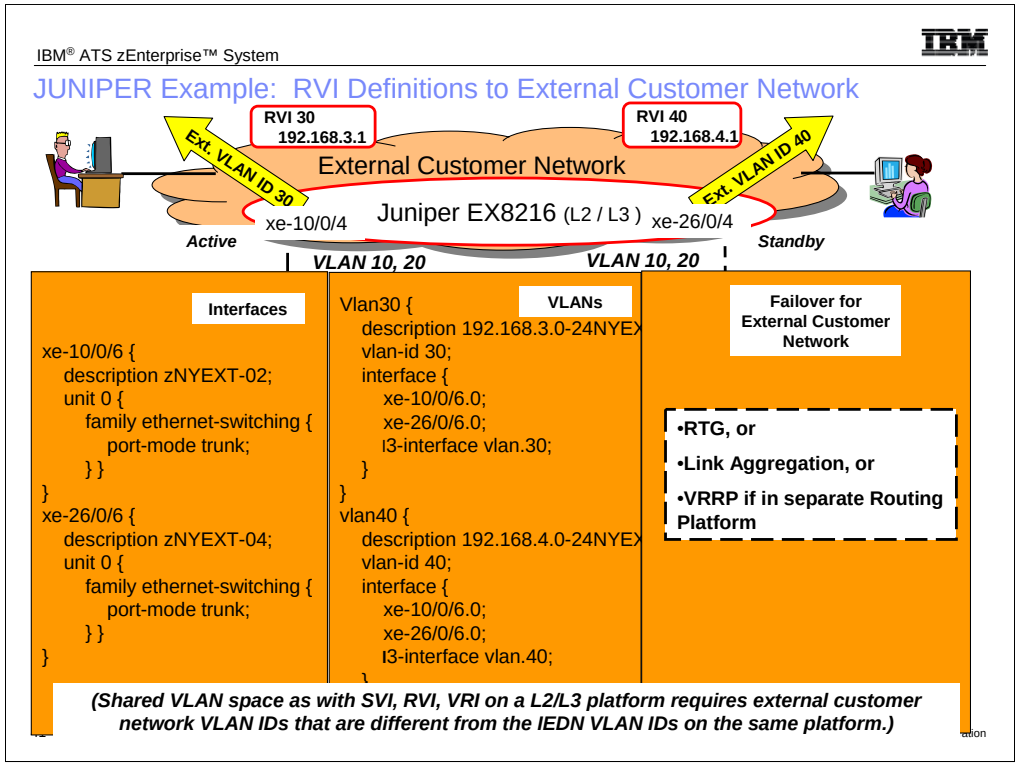
Great care must be taken to create an IEDN domain that is separate from any of the layer 2 domains in the external network.This puts an extra burden on the VLAN designer and the administrator of the switch.

TESTING RESULTS: Cisco's Virtual Chassis implementation has been tested with this configuration by a customer using SVI and FlexLink. IBM has tested a similar configuration using Juniper's RVI implementation together with Redundant Trunk Group.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

Possible Link Failover Technologies:

•HotLink = IBM Systems Networking (formerly BNT) ; FlexLink = Cisco ; Protected Link Group = Brocade; Redundant Trunk Group (RTG) = Juniper



Portions of the definition have been elided for purposes of the visual. Virtual Interface implementations (also known as Switched Virtual Interfaces – SVI – or Routed Virtual Interfaces – RVI – or Virtual Routed Interfaces – VRI) on Layer 2/Layer 3 Switches -- together with a failover technology like HotLink, FlexLink, Protected Link Group – are supported, but with caution. **Each active switch port on a L2/L3 platform implemented with routed interfaces using SVI, RVI, or VRI must be associated with a unique VLAN ID. Therefore, with connections of the same VLAN ID into two different TOR ports, you must implement a link failover technology that leaves only one of the two connections active and the other in standby mode. (Otherwise you would have two active ports with the same VLAN ID in the attached platform and cause a loop.)**

Great care must be taken to create an IEDN domain that is separate from any of the layer 2 domains in the external network.This puts an extra burden on the VLAN designer and the administrator of the switch.

TESTING RESULTS: Cisco's Virtual Chassis implementation has been tested with this configuration by a customer using SVI and FlexLink. IBM has tested a similar configuration using Juniper's RVI implementation together with Redundant Trunk Group.

Virtual Interface Implementations: SVI (Cisco) – Switched Virtual Interfaces; RVI (Juniper) – Routed Virtual Interfaces; VRI (Brocade) – Virtual Routed Interfaces

Possible Link Failover Technologies:

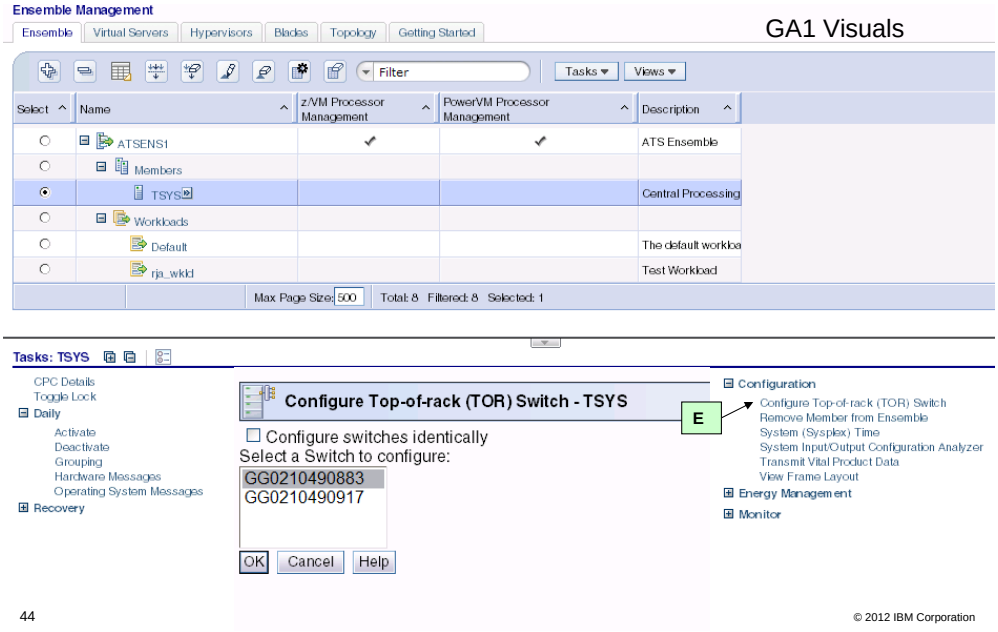
•HotLink = IBM Systems Networking (formerly BNT) ; FlexLink = Cisco ; Protected Link Group = Brocade; Redundant Trunk Group (RTG) = Juniper

Appendix:
VLAN Modes and Configuring the Top-of-Rack Switches with Unified Resource Manager

Access vs. Trunk Mode for External Connectivity to the TOR

- **General Support Guidelines:**
  - **The IEDN VLAN must terminate at a ROUTED endpoint on the external platform adjacent to the zBX.**
    - If TOR is configured in **ACCESS Mode** ...
      - The TOR tags the packets with the proper IEDN VLAN ID.
    - If TOR is in **TRUNK Mode** ...
      - The TOR validates that the packets are tagged with the proper IEDN VLAN ID.

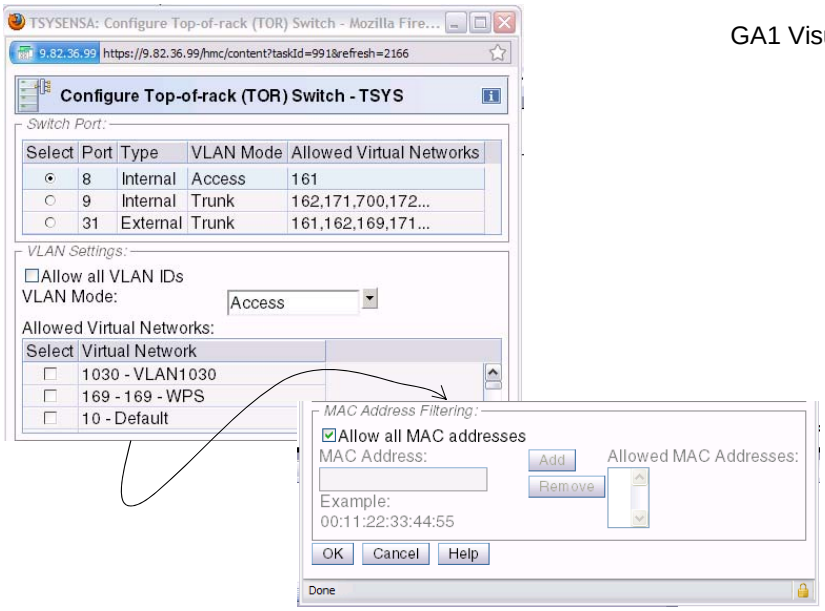
Top-of-Rack (TOR) Switch at the TSYS Node (Member)



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. Alternatively, you can check the box that allows you to configure a single switch but propagate the same definitions to the second switch: “Configure switches identically.”

Configuring the TOR Switch

GA1 Visuals



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.

End of Topic