A CMI Technical White Paper:

PowerVM Virtual Ethernet

By Lance Hawkes, Sr. Solutions Architect, CMI

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

As processor speeds have increased and multiple cores have become consolidated onto a single processor virtualization or the ability to consolidate multiple workloads on a single system has become necessary. Server virtualization improves utilization rates of CPU, memory and I/O resources making better use of IT assets. A virtualized environment is more responsive to business needs through dynamic re-allocation of resources to meet those needs.

PowerVM is the virtualization solution for AIX, IBM i, and Linux operating systems on the IBM POWER technology. While PowerVM offer many industry leading technologies, the purpose of this paper is to discuss the various options available to virtualize Ethernet on the IBM POWER platform.

Virtualized Ethernet and Benefits

Ethernet enables in-memory connections represented as virtual Ethernet adapters between LPARs or virtual machines. These connections have similar characteristics to physical high-bandwidth Ethernet adapters and support the same industry standard protocols.

Virtualizing an Ethernet adapter allows for a physical Ethernet adapter to be shared among multiple virtual servers/LPARs that increases utilization rates of the physical adapter. This is especially important as Ethernet adapters become faster and denser with multiple ports. Higher utilization rates derive greater value from the physical resource. Implementation procedures can be found in the IBM Redbook IBM PowerVM Virtualization Introduction and Configuration

Before we begin to explain the different methodologies, it would be best to explain some of the common terms used in PowerVM virtual networking:

Virtual Ethernet*	The collective name for technologies that comprise a virtualized Ethernet environment.
Virtual Ethernet Adapter	A hypervisor-provided network adapter that is configured within a partition to enable communication in a virtual Ethernet.
Virtual LAN	Technology pertaining to the segmentation of virtual Ethernet networks. More commonly referred to as VLANs.
Virtual Switch	An in-memory, hypervisor implementation of a Layer-2 switch that supports tagged and untagged ports.

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

Shared Ethernet Adapter	A Virtual I/O Server software adapter that bridges physical and virtual Ethernet networks.
Integrated Virtual Ethernet	The collective name for the hypervisor-attached, physical Ethernet port and its capability to be shared among partitions.
Host Ethernet Adapter	A hypervisor-provided network adapter that is backed by a physical Ethernet port in an IVE configuration.

* Terms taken from PowerVM Virtualization Introduction and Configuration Redbook p.144

It is important to note that additional tuning is required to derive the highest possible value from these virtualized resources. Tuning recommendations can be found in IBM PowerVM Virtualization Managing and Monitoring.

There are times, however, that it is not advisable to virtualize Ethernet resources. If you have servers that are already fully utilizing the entire bandwidth of an Ethernet adapter, it is not advisable to use virtualized Ethernet. Backup servers that utilize the LAN for backup purposes would be best served by dedicated physical adapters.

Four Different Approaches at Work

There are four primary methods of virtualizing Ethernet adapters in a PowerVM environment.

In each of these cases we are assuming that the environment will consist of two Virtual I/O Servers for higher redundancy and resiliency.

1. Single Virtual Switch NIB (Network Interface Backup)

By default the Power hypervisor comes with a single virtual switch already configured called ETHERNET0. This configuration uses the default virtual switch and each VIO server is defined with a trunk port on a different VLAN. The client is then created with two virtual Ethernet adapters ent0 and ent1. Using Etherchannel on the client you designate which of the virtual Ethernet adapters will the primary network adapter and select the other as a backup. This Etherchannel configuration is known as Network Interface Backup or NIB. The Etherchannel process then creates an additional network adapter that you apply an interface to an assign an IP address as shown in the following illustration.

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

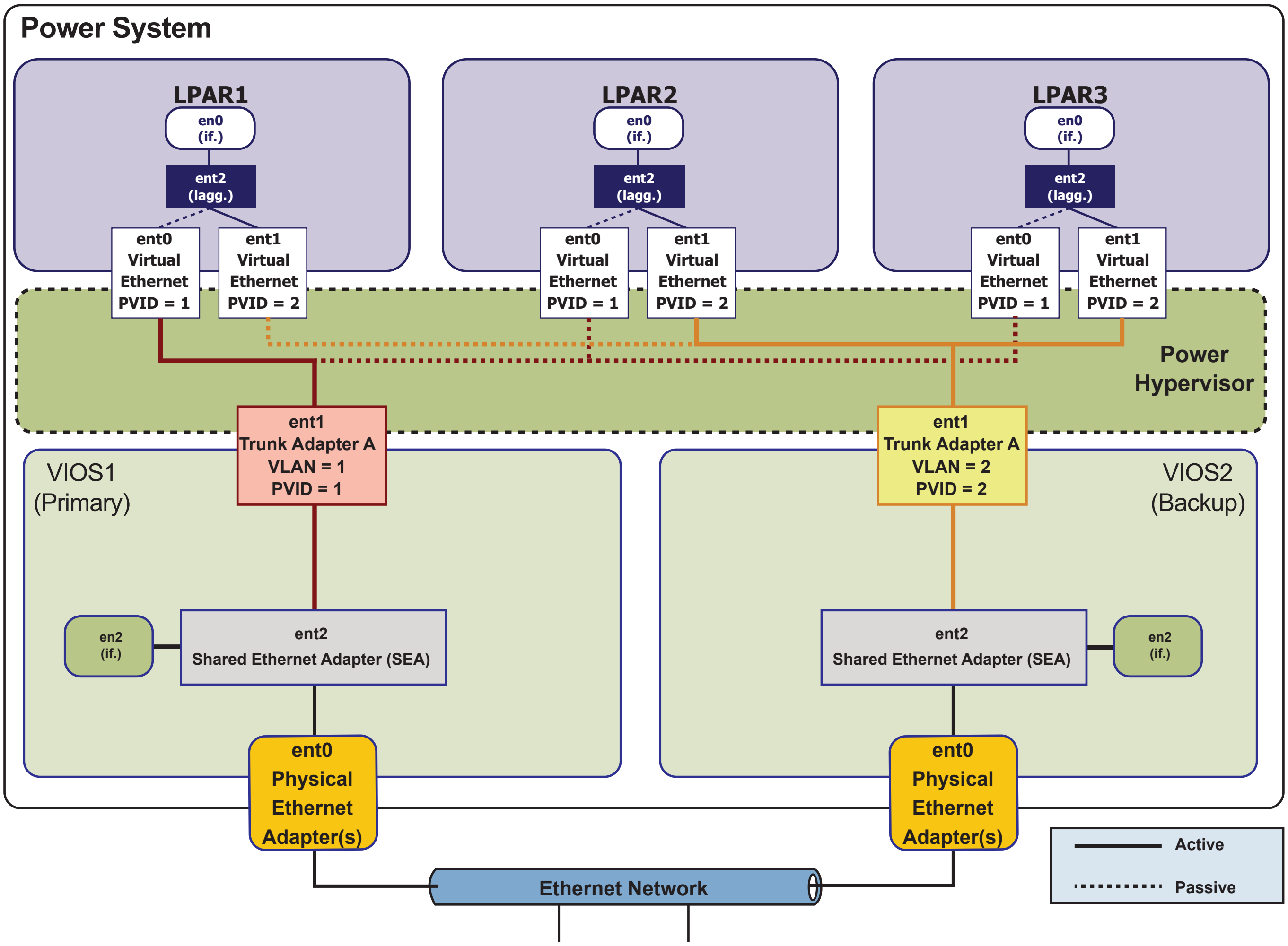

Figure 1: Classic NIB

This configuration allows the administrator to manually select the primary path for network traffic and is then able to balance that traffic between both virtual I/O servers ensuring that all physical network adapters are being utilized.

This is a very simple configuration but allows for all network resources to be utilized. The downside of this configuration is that due to the single virtual switch and each client being presented with virtual adapters on different VLAN's no further VLAN tagging is possible. If VLAN tagging is a requirement in your environment, this configuration will not work. Because of the relative simplicity and advanced functionality of some of the following technologies we will discuss, this configuration while valid is rarely implemented any longer.

2. Shared Ethernet Adapter Failover (SEAF)

Shared Ethernet Adapter Failover was one of the earlier options available for virtual networking. Unlike the classic NIB model, it allows for VLAN tagging and is very clean to the client as the client is presented with a single virtual Ethernet adapter. As shown in the illustration below, one or more trunk adapters is defined on each VIO server with one of the VIO servers designated as the primary path for all network traffic as indicated by the priority. All of the clients then use that single VIO server for all network traffic unless there is a network interruption identified by the control channel that is created for path fault detection between the VIO servers. Once a fault is detected, backup VIO server then becomes the primary for all network traffic. The administrator is also able to manually force a failover between the VIO servers, but the primary disadvantage of this configuration is that all network traffic is forced through a single VIO server. This does not allow for full utilization of the network resources and may require additional network resources if the client network load is substantial. This is probably the best known and most widely implemented option for virtual networking. The IBM Redbooks PowerVM Introduction and Configuration and PowerVM Managing and Monitoring write extensively and provide step-by-step instructions for implementing this model. This model is also becoming dated and additional features have been added that will further decrease future implementations of this model.

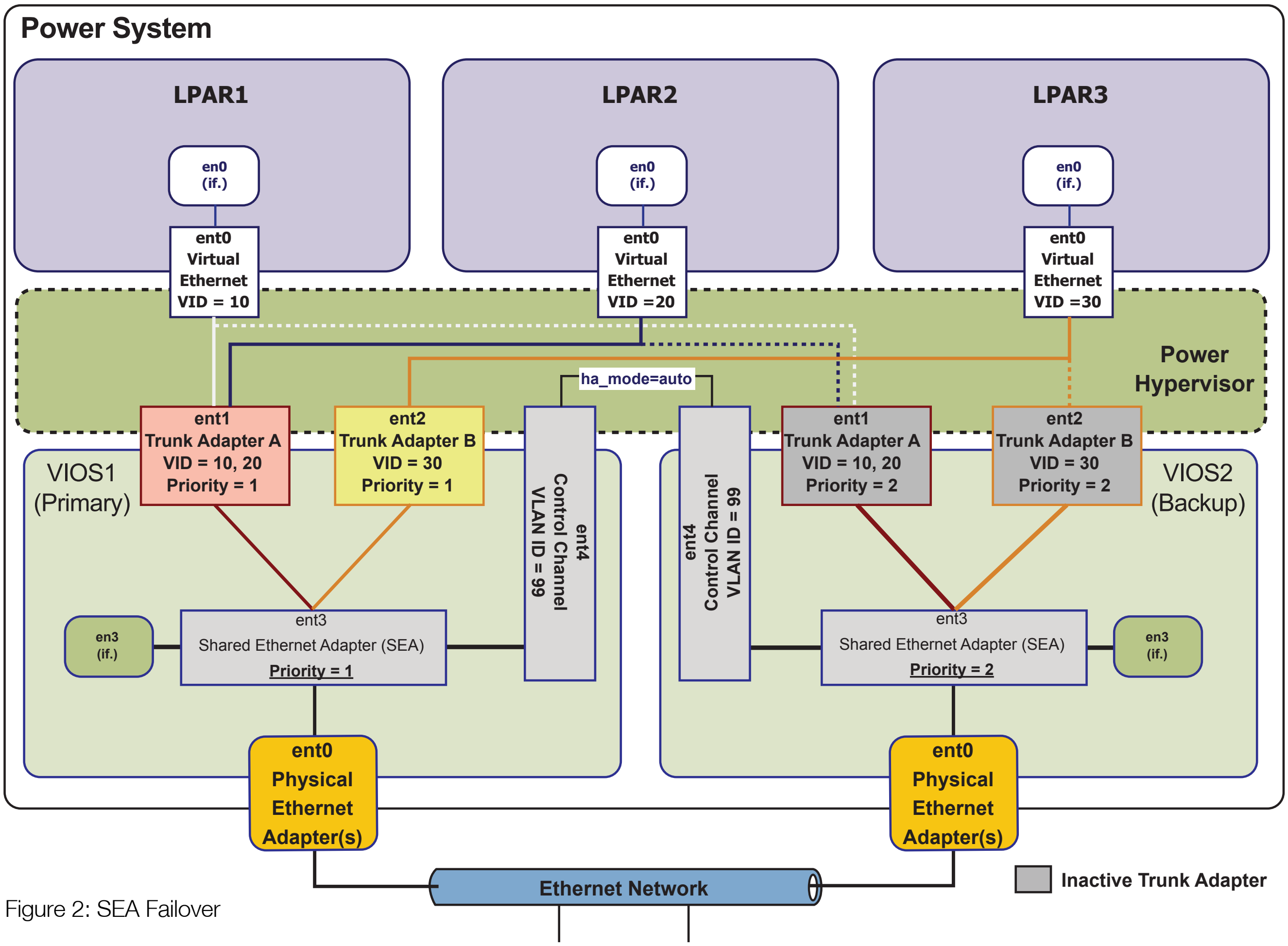

Figure 2: SEA Failover

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

3. Shared Ethernet Adapter Failover with Load Sharing

Shared Ethernet Adapter Failover with Load Sharing is nearly identical to the classic Shared Ethernet Adapter Failover model with some very slight changes. When creating the Shared Ethernet Adapter and control channel instead of specifying ha_mode as auto and creating a single trunk port, you must create multiple trunk ports and specify sharing as the ha_mode as shown in the following illustration.

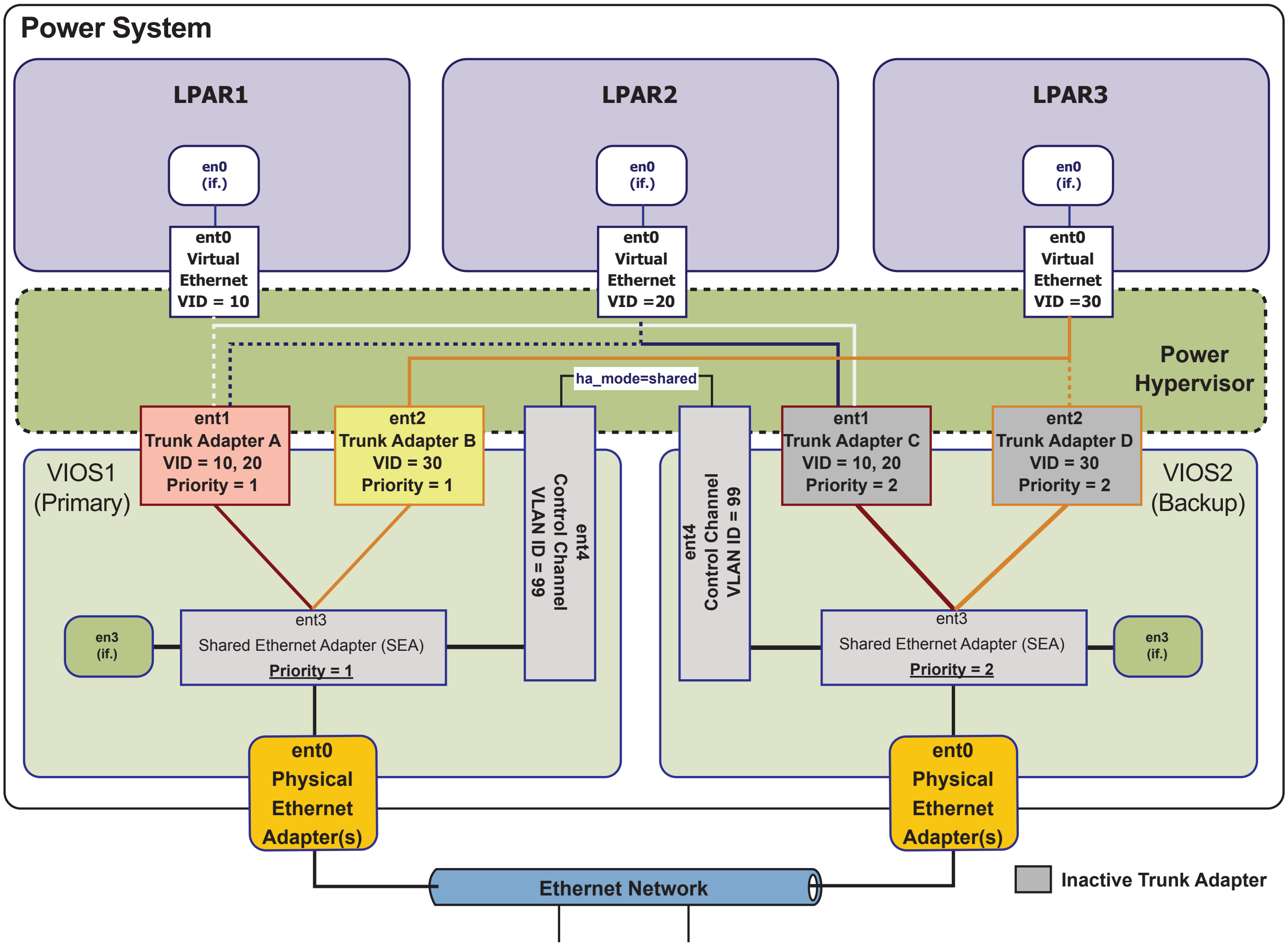

Figure 3: Shared Ethernet Adapter Failover with Load Sharing

This model is only available on VIO servers running version 2.2.1.0, or later. It also requires that two or more trunk adapters are configured for the primary and backup SEA pair. Load sharing mode must be enabled on both the primary and the backup. The VLAN definitions of the trunk adapters must also be identical between the SEA pairs. Configuration instructions are provided in the IBM PowerVM Virtualization Managing and Monitoring Redbook beginning on page 138.

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

As the name implies, the clear advantage of this model is that network traffic can be shared by both VIO servers allowing all network resources to be used. However, if VLAN tagging is not implemented in your environment, this model cannot be implemented, as it requires multiple trunk adapters on separate VLANs. The load sharing algorithm determines which trunk adapters will be activated. You cannot specify the active trunk adapters manually in the load sharing mode. Therefore, if you have multiple VLANs but the traffic between the VLANs is not well balanced, load sharing will not assist with that balancing as all traffic for a particular VLAN must go out the same VIO server. If you have a particular LPAR with a large network load, you are unable to manually determine which VIO server will service the load. However, if you have many VLANs that have similar network traffic profiles, this model will manage network load sharing and allow you to utilize all of your network resources.

4. Multiple Virtual Switches Methodology with NIB

Beginning with the Power6 processor multiple virtual switch technology was introduced. This virtual networking implementation model is very similar to the classic NIB model, but with the advent of multiple virtual switches, VLAN tagging became supported. As is illustrated in the following diagram the client LPARs are presented with two virtual Ethernet adapters, however, unlike the classic NIB model each adapter is logically connected to a different virtual switch. Using network interface backup through the Etherchannel configuration on the client one network adapter is designated as the primary and the other as backup. Because only one port is active a loop is not created on the network and spanning tree protocol will not shut down the ports. Therefore there can be multiple adapters in the same VLAN and VLAN tagging can be permitted because MAC address stability is guaranteed.

The greatest benefit of this model is that the network resources in both VIO servers can be fully utilized at the same time ensuring that the hardware investment is realized. The administrator determines which path will be primary and therefore one with understating ofthe business/application needs can also make sure SLA's are met by manually balancing network load.

The following diagram illustrates this model.

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

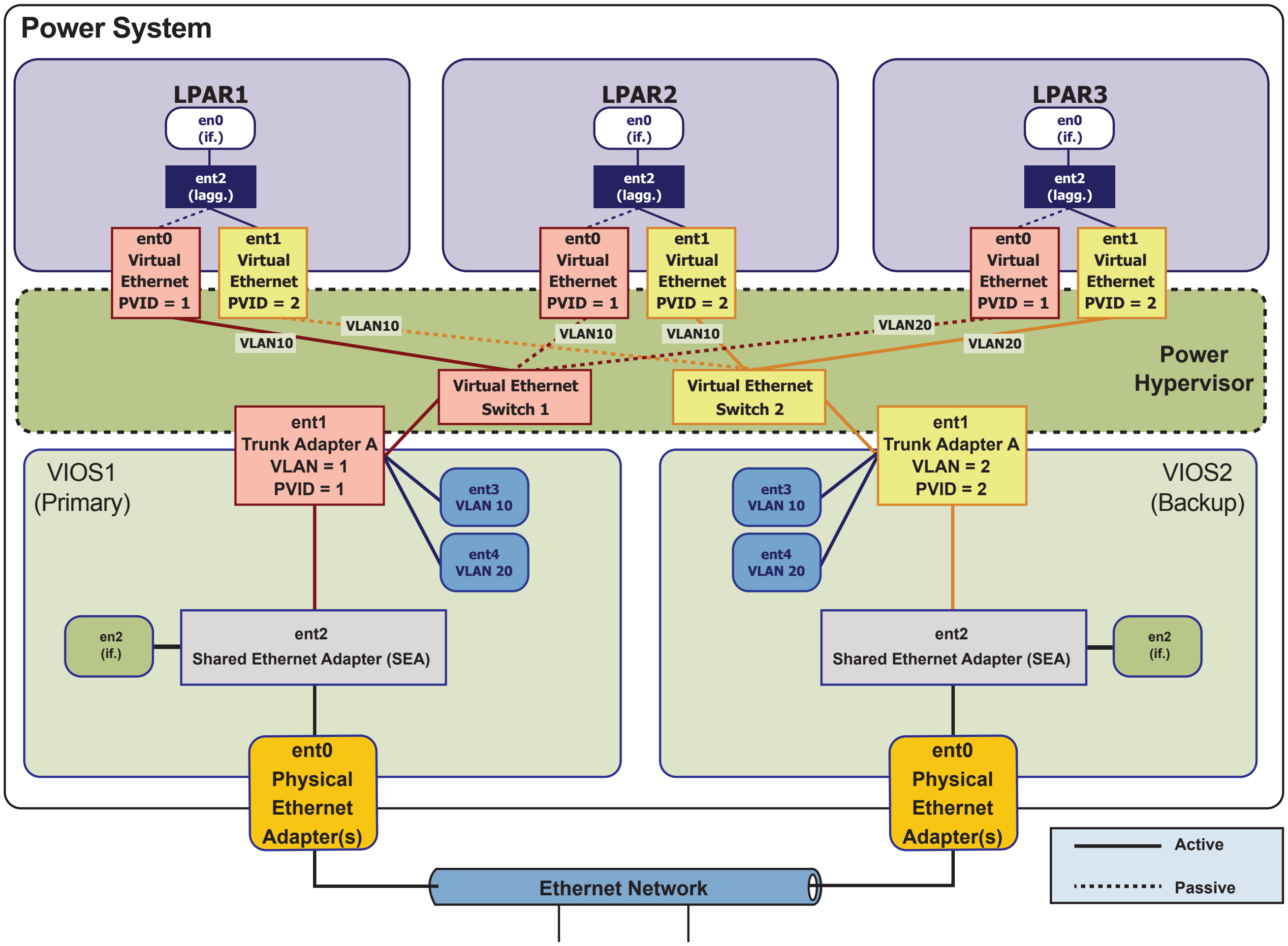

Figure 4: Multiple Virtual Switches and NIB

5. Optional Note: Port Virtualization with LHEA (Logical Host Ethernet Adapter) – Why it's being phased out.

When Power6 was introduced the systems were created with an Integrated Virtualized Ethernet adapter. This was a physical Ethernet adapter backed by the Power hypervisor that allowed the individual ports to be virtualized outside of the VIO server. The physical port itself was referred to as the logical host Ethernet adapter coupled with the software components and the POWER hypervisor is known as the Integrated Virtualized Ethernet adapter. This is detailed in the IBM PowerVM Virtualization Introduction and Configuration Redbook on page 169. The IVE is an ultra-high performance Ethernet adapter connected directly to the GX+ bus of the server and

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

can communicate directly with the logical partitions running on the system. The IVE acts as layer 2 switch and allows the capability to virtualize the individual ports to be used by multiple partitions/virtual machines. Each logical port then receives as unique MAC address.

This technology has greater performance possibilities than any of the technologies that use a Shared Ethernet Adapter as that virtualization layer requires more CPU and memory resources to operate than the IVE. With the MMC class of Power7 servers, the IVE is no longer available. There were two major disadvantages of using the IVE. First, the IVE is not a hot-swappable adapter, therefore the failure of this adapter requires that the entire server and all its logical partitions be shutdown for replacement. This does not preclude the use of a spare network adapter in the event of a hardware failure, but it is something to consider. Secondly, logical partitions using the IVE cannot be migrated with logical partition mobility. Some have worked around this by temporarily removing the IVE from the partition configuration for the move and then adding it back again, but this requires considerable manual effort. It seems clear however that IBM is moving away from this technology.

Minimum Requirements Ahead

With the exception of the IVE, in order to use any of these technologies, you must have multiple virtual I/O servers. Best practices would dictate that you have at least two VIO servers simply for maintenance and availability purposes.

You will need at least two Ethernet adapters, one for each VIO server though four adapters would be preferable for availability purposes.

The logical partitions must be running AIX 5.3 or above and for Shared Ethernet Adapter Failover with Load-Sharing the VIO server must be running version 2.2.1.0 or above which would also require that the server itself be either Power6 or higher.

Classic NIB and Shared Ethernet Adapter Failover are supported on Power5 technology though all Power5 hardware has been withdrawn from standard support.

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com

Conclusion

IBM has been at the forefront of virtualization technologies and the different options of virtualized Ethernet show the progression and inclusion of advanced features and functions with Shared Ethernet Adapter Failover with Load Sharing and multiple virtual switches with NIB being the latest iteration. CMI believes that for most implementations multiple virtual switches is the best option providing the greatest flexibility and availability among the different options. It is also the only option that provides manual network path designation with VLAN tagging allowing an administrator to make intelligent choices for balancing network traffic and utilizing resources.

CMI (Chouinard & Myhre, Inc.)
655 Redwood Hwy, Ste. 102
Mill Valley, CA 94941
415.480.3636
www.cm-inc.com