HOWTO: Implement SEA Failover with Dual VIOS

Contributed by Michael Felt

In 2004 IBM released POWER5 with a very "Power"ful feature built-in - Virtual Ethernet. An additional feature known as Shared Ethernet Adapter (SEA) enables client partitions to communicate with via IP protocols without a requirement for a physical ethernet adapter. In 2005 IBM released an updated version of the virtual I/O server (VIOS)that could be used to eliminate the VIOS as a SPOF - SEA Failover.

This article presents the evolution of SEA configurations while showing you how to configure SEA Failover according to today's best practice.

Virtual Ethernet is in all POWER5 and newer systems

Back in 2004 IBM released POWER5 with a very "Power"ful feature built-in - Virtual Ethernet.
Using Virtual Ethernet partitions (also known as LPAR's) could communicate with each other without a physical ethernet adapter. However, many servers need to communicate with servers outside a managed system.

One simple solution was/is to assign a physical adapter to one partition as well as a virtual ethernet adapter.
If both adapters, physical and virtual, were given IP addresses in different IP networks this partition could function as a router. As a router it did not matter if the partition ran AIX or Linux (on Power).
Since Virtual Ethernet is included in all POWER5 and newer POWER systems any partition can communicate with the external ethernet as long as at least one partition has a physical interface.

PowerVM

The POWER Hypervisor (PHYP) is part of the POWER firmware and provides (nearly) all the functions of an enterprise ethernet switch. But this switch is local - it only functions within the managed system. For partitions to be able to act as if they have a direct connection to the physical ethernet infrastructure an addtional feature is needed. This feature is called Shared Ethernet Adapter (SEA). This feature is supported by a specialized partion known as the Virtual I/O Server (VIOS). The other feature that has always been supporte by the VIOS is Virtual SCSI (vSCSI). Currently the VIOS also supports AMS and NPIV (Active Memory Sharing and N_port ID Virtualization, respectively).

When POWER5 was first released in 2004 the PHYP could support only one so-called trunk adapter. If you wanted dual ethernet support in the client you were required to create two virtual servers, each with a SEA defined on a seperate PVID (or PHYP VLAN) and configure a etherchannel interface (also know as NIB - Network Interface Backup) in the client.
This solution is still quite common. However, since about May 2005 POWER firmware and VIOS support a feature known as SEA Failover. In most cases this solution is better than NIB in the client.

The rest of this article shows how so-called SEA best practice has evolved into the current best practice and shows how to (re) configure systems to use SEA Failover.

The starting point is a number of clients with no physical ethernet adapters connected via the PHYP to a partition with a physical adapter. As mentioned above this partition could be a router that performs ip_forwarding.

SEA is preferred

SEA is preferred for many reasons. The primary reason is ease of use. Using SEA the client sees and can be seen as network managers intend. The
drawback to using a router is twofold: it requires extra processing power and clients are invisible to network management tools. When a VIOS is used to implement a SEA the virtual and physical adapters are bridged. This
function is called Shared Ethernet Adapter, also known as SEA.

Diagram 1 shows the intial setup of a VIOS partiton after it's first boot. At this point all configuration must be performed via the HMC (Integrated Virtualization Manager (IVM), an additional feature
of the VIOS, is not discussed here because IVM does not support SEA Failure).
From this situation only one command is needed to permit the clients to communicate with external systems:

```
$ mkvdev -sea ent0 -vadapter ent1 -default ent1 -defaultID 1
```

This command creates a new interface - the SEA. This interface can be given an IP address (using the padmin version of mktcpip, or switching to a root prompt using oem_setup_env. Once the SEA has been given an IP address the VIOS can communicate with the external network (e.g. HMC for DLPAR commands, or access via telnet/ssh
by system administrators).

Gradually people recognized that having an IP address on the SEA interface was not always desired. In any case, having an
address on the SEA is not a requirement for it to function. As SEA became more common
a new best practice was adopted for VIOS. Rathar than only one virtual ethernet adapter - two virtual adapters were defined. The first one was still used for the SEA bridging function while the second one was used for assigning the VIOS IP address. The advantage is that the VIOS can have an IP address independent of the status of an SEA. Without a SEA defined a VIOS could be configured as follows.
Before removing the SEA use DPLAR command to create a new virtual interface (and don't forget to add it to the current profile, or better - save the profile after adding the interface. Do NOT run cfgmgr right away.
Before you do that, remove the SEA interface as padmin using:

```
$ rmdev -dev ent2

$ oem_setup_env

# cfgmgr
```

(If DLPAR is not working, remove the SEA, update the profile and reboot.)

```
# chdev -l en2 -a netaddr=A.B.C.D -a netmask=255.255.255.0 -a state=up

# exit
```

The IP address that was on the SEA is now being assigned to the new virtual IP interface (en2). Assuming the clients and VIO1 are inthe same IP network they should be able to ping each other without the SEA being defined.
Use the same command as above to create the SEA - ent3 is now defined and both the VIO and the clients have external access
via the SEA interface.

```
$ mkdev -sea ent0 -vadapter ent1 -default ent1 -defaultid 1
```

Now the VIOS is configured similiar to a client in that it's IP configuration
is independent of the status (defined, undefined) of the SEA.

The next general practice was to create a hardware etherchannel (or NIB) of the physical adapters of the VIOS partition besides the two virtual ethernet adapters. The recommendation is to remove all ethernet defintions from the ODM and then rerun cfgmgr. This means that the virtual interface used by the VIOS for it's IP address will change from en2 to en3.

Use DLPAR to add physical ethernet port before removing IP address!!

```
$ rmdev -dev ent3 # remove SEA

$ rmdev -dev en2 # remove tcpip definition

$ rmdev -dev ent2 #remove virtual ethernet adapters

$ rmdev -dev ent1

$ oem_setup_env
```

```
# cfgmgr

# chdev -l en3 -a netaddr=A.B.C.D -a netmask=255.255.255.0 -a state=up

# exit
```

Again, from this configuration the VIOS and clients can communicate via the PHYP virtual ethernet.
To increase the availability of the SEA we first create a LAN aggregate (etherchannel) over the physical ethernet interfaces. As padmin:

$ mkvdev -lnagg ent0 ent1

This creates the interface ent4.

Now to create the SEA we use ent4 as the physical interface rather than ent0 as before. Note, since there are two physical interfaces the number of the trunk interface has changed as well. Our command to create the SEA is now as follows:

$ mkvdev -sea ent4 -vadapter ent2 -default ent2 -defaultid 1

And the SEA is created as ent5

This is a resilient implementation of the SEA. However, the VIOS itself remains a SPOF (single point of failure).

Many administrators choose to implement an identical VIOS except that a different PVID (or VLAN ID) is used for the virtual ethernet. Fortunately the physical interfaces can (actually must) be in the same VLAN.

The main reason
this (see above) is not the favored best practice is because all existing clients must be reconfigured. For this to work their IP address that was assigned to en0 must be reassigned to the client based NIB adapter (en2 and ent2) built on the virtual adapters ent0 and ent1 (connecting to VIO1 and VIO2 respectiveally)

Best Practice - SEA Failover setup

Current best practice is dual VIOS support and SEA Failover. The starting point is two physical ports that will be set into a lan aggregate and three
virtual ethernet adapters. When building on a production VIOS the initial setup will look something like this.

In the new VIOS the third virtual ethernet adapter (veth) has been placed in a different VLAN. This veth will be used for control messages. One of the other virtual ethernet adapters (here the first one) is set to have a different priority (e.g., second) than the first (already operational in VIO1) virtual ethernet adapter. (If you forget to give it a different value the partition will not start.) The VIOS (e.g., VIO2) IP address can be assigned to the remaining interface.

The lan aggregate is made the same as before. For the failover ability we need some additional arguments when creating the SEA adapter. On VIO2 execute the following commands:

$ mkvdev -lnagg ent0 ent1

$ mkvdev -sea ent5 -vadapter ent2 -attr ha_mode=auto ctl_chan=ent4

And now on VIO1 you can remove the SEA and lnagg interfaces

$ rmdev -dev ent5

$ rmdev -dev ent4

Since the en3 interface is now using the SEA provided by VIO2 you can use DLPAR actions to create a new virtual ethernet interface (ent4) and then use cfgmgr to get the interface activated. Do not forget to save the profile after creating the interface using DLPAR or you will may not have a proper profile.

Make sure that
this new interface has the same PVID as the control interface in VIO2.
Now your combined setup should look something like this:

Finishing touches

Now we can repeat the commands we used on VIO2 earlier and VIO1 will again become the primary SEA.
Now all the client partitions have the benifit of resilient VIOS SEA configuration with automatic
SEA failover to the other VIOS (i.e. no SPOF) without making any changes to the client partition configuration.

$ mkvdev -sea ent5 -vadapter ent2 -attr ha_mode=auto ctl_chan=ent4