virtual ports correspond directly to virtual Ethernet adapters that can be assigned to partitions from the HMC or IVM. There is no need to explicitly attach a virtual Ethernet adapter to a virtual Ethernet switch port. To draw on the analogy of physical Ethernet switches, a virtual Ethernet switch port is configured when you configure the virtual Ethernet adapter on the HMC or IVM.

For AIX, a virtual Ethernet adapter is not much different from a physical Ethernet adapter. It can be used:

- To configure an Ethernet interface with an IP address onto it
- To configure VLAN adapters (one per VID) onto it
- As a member of a Network Interface Backup adapter

But it cannot be used for EtherChannel or Link Aggregation.

The POWER Hypervisor’s virtual Ethernet switch can support virtual Ethernet frames of up to 65408 bytes size, which is much larger than what physical switches support: 1522 bytes is standard and 9000 bytes are supported with Gigabit Ethernet Jumbo Frames. Thus, with the POWER Hypervisor’s virtual Ethernet, you can increase TCP/IP’s MTU size to 65394 (= 65408 - 14 for the header, no CRC) in the non-VLAN-case and to 65390 (= 65408 - 14 - 4 for the VLAN, again no CRC) if you use VLAN.

Increasing the MTU size could benefit performance because it may improve the efficiency of the transport. This is dependant on the communication data requirements of the running workload.

### 2.7.3 Sharing physical Ethernet adapters

Virtual Ethernet enables inter-partition communication on the same server. There are two approaches to connect a virtual Ethernet to an external network:

**Routing**

- Layer-3 IP packet forwarding

**Bridging**

- Layer-2 Ethernet frame forwarding

**Routing**

By enabling the IP forwarding capabilities of an AIX or Linux partition with virtual and physical Ethernet adapters, the partition can act as router. Figure 2-30 on page 94 shows a sample configuration. The client partitions have their default routes set to the partition. This routes the traffic to the external network.

**Note:** In this type of configuration, the partition that routes the traffic to the external network cannot be the Virtual I/O Server. This is because you cannot enable IP forwarding from the Virtual I/O Server command line interface.
The routing approach has the following characteristics:

- It does not require the purchase of either PowerVM features and use of a Virtual I/O Server.
- IP filtering, firewalling, or Quality of Service (QoS) could be implemented on these routing partitions.
- The routing partitions could also act as endpoints for IPsec tunnels, thus providing for encrypted communication over external networks for all partitions, without having to configure IPSec on all partitions.
- Continuous availability can be enhanced by implementing more than one such routing partition and by configuring IP multipathing on the clients, or by implementing IP address fail over on routing partitions. This is discussed in 4.3, “Shared Ethernet redundancy” on page 254.

**Shared Ethernet Adapter**

A Shared Ethernet Adapter (SEA) can be used to connect a physical Ethernet network to a virtual Ethernet network. It also provides the ability for several client partitions to share one physical adapter. Using a SEA, you can connect internal...
and external VLANs using a physical adapter. The SEA hosted in the Virtual I/O Server acts as a layer-2 bridge between the internal and external network.

A SEA is a layer-2 network bridge to securely transport network traffic between virtual Ethernet networks and physical network adapters. The Shared Ethernet Adapter service runs in the Virtual I/O Server. It cannot be run in a general purpose AIX or Linux partition.

**Tip:** A Linux partition can provide bridging function as well, with the `brctl` command.

There are some things to consider on the use of SEA:

- Virtual Ethernet requires the POWER Hypervisor and PowerVM feature (Standard or Enterprise Edition) and the installation of a Virtual I/O Server.
- Virtual Ethernet cannot be used prior to AIX 5L Version 5.3. Thus, an AIX 5L Version 5.2 partition will need a physical Ethernet adapter.

The Shared Ethernet Adapter allows partitions to communicate outside the system without having to dedicate a physical I/O slot and a physical network adapter to a client partition. The Shared Ethernet Adapter has the following characteristics:

- Virtual Ethernet MAC addresses of virtual Ethernet adapters are visible to outside systems (using the `arp -a` command).
- Unicast, broadcast, and multicast is supported, so protocols that rely on broadcast or multicast, such as Address Resolution Protocol (ARP), Dynamic Host Configuration Protocol (DHCP), Boot Protocol (BOOTP), and Neighbor Discovery Protocol (NDP) can work across an SEA.

In order to bridge network traffic between the virtual Ethernet and external networks, the Virtual I/O Server has to be configured with at least one physical Ethernet adapter. One SEA can be shared by multiple virtual Ethernet adapters and each can support multiple VLANs. Figure 2-31 on page 96 shows a configuration example of an SEA with one physical and two virtual Ethernet adapters. An SEA can include up to 16 virtual Ethernet adapters on the Virtual I/O Server that share the physical access.
A virtual Ethernet adapter connected to the Shared Ethernet Adapter must have the Access External Networks check box (named the trunk flag in some earlier releases of the HMC) enabled. Once an Ethernet frame is sent from a virtual Ethernet adapter on a client partition to the POWER Hypervisor, the POWER Hypervisor searches for the destination MAC address within the VLAN. If no such MAC address exists within the VLAN, it forwards the frame to the virtual Ethernet adapter on the VLAN that has the Access External Networks option enabled. This virtual Ethernet adapter corresponds to a port of a layer-2 bridge, while the physical Ethernet adapter constitutes another port of the same bridge.
The SEA directs packets based on the VLAN ID tags. One of the virtual adapters in the Shared Ethernet Adapter on the Virtual I/O Server must be designated as the default PVID adapter. Ethernet frames without any VLAN ID tags that the SEA receives from the external network are forwarded to this adapter and assigned the default PVID. In Figure 2-31 on page 96, ent2 is designated as the default adapter, so all untagged frames received by ent0 from the external network will be forwarded to ent2. Since ent1 is not the default PVID adapter, only VID=2 will be used on this adapter, and the PVID=99 of ent1 is not important. It could be set to any unused VLAN ID. Alternatively, ent1 and ent2 could also be merged into a single virtual adapter ent1 with PVID=1 and VID=2, being flagged as the default adapter.

When the SEA receives or sends IP (IPv4 or IPv6) packets that are larger than the MTU of the adapter that the packet is forwarded through, either IP fragmentation is performed, or an ICMP packet too big message is returned to the sender, if the Do not fragment flag is set in the IP header. This is used, for example, with Path MTU discovery.

Theoretically, one adapter can act as the only contact with external networks for all client partitions. For more demanding network traffic scenarios (large number of client partitions or heavy network usage), it is important to adjust the throughput capabilities of the physical Ethernet configuration to accommodate the demand.

There are several different ways to configure physical and virtual Ethernet adapters into Shared Ethernet Adapters to maximize throughput:

- Using Link Aggregation (EtherChannel), several physical network adapters can be aggregated. Refer to 4.3.4, “Using Link Aggregation on the Virtual I/O Server” on page 261 for more details.
- Using several Shared Ethernet Adapters provides more queues and more performance.
Other aspects that have to be taken into consideration are availability (refer to 4.3, “Shared Ethernet redundancy” on page 254) and the ability to connect to different networks.

**When to use routing or bridging**
Where several servers are consolidated onto a single system or where LPARs are moved to another server, bridging is often the preferred choice. This is because the network topology does not have to be changed and IP subnets and IP addresses of the consolidated servers can stay unmodified. Even an existing multiple VLAN scheme can be bridged.

Routing may be worth consideration under various circumstances:

- Additional functions to basic packet forwarding are required and should be performed in a central place:
  - IP filtering
  - Firewalling
  - QoS Routing
  - IPsec tunneling

- The external network is a layer-3-switched Ethernet with the dynamic routing protocol OSPF, as found in many IBM System z™ environments.

- Avoid purchasing the PowerVM feature (Standard or Enterprise Edition) because the routing approach does not require the use of the Virtual I/O Server.

To summarize: in most typical environments, bridging will be the most appropriate option, being simpler to configure. It should be considered as the default approach.

### 2.7.4 Virtual and Shared Ethernet configuration example

After having introduced the basic concepts of VLANs, virtual Ethernet, and Shared Ethernet Adapters in the previous sections, this section discusses in more detail how communication between partitions and with external networks operates. The sample configuration in Figure 2-32 on page 99 is used as an example.

The configuration is using four client partitions (Partition 1 through Partition 4) running AIX and one Virtual I/O Server (VIOS). Each of the client partitions is defined with one virtual Ethernet adapter. The Virtual I/O Server has a Shared Ethernet Adapter (SEA) that bridges traffic to the external network.
Inter-partition networking

Partition 2 and Partition 4 are using virtual Ethernet adapters with the Port virtual LAN ID (PVID) only. This means that:

- The operating system running in such a partition is not aware of the VLANs.
- Only packets for the VLAN specified as PVID are received.
- Packets have their VLAN tag removed by the POWER Hypervisor before the partitions receive them.
- Packets sent by these partitions have a VLAN tag attached for the VLAN specified as PVID by the POWER Hypervisor.

In addition to the PVID, the virtual Ethernet adapters in Partition 1 and Partition 3 are also configured for VLAN 10 using a VLAN Ethernet adapter (ent1) and network interface (en1) created through the `smitty vlan` command on AIX (using the `vconfig` command on Linux). This means that:

- Packets sent through network interfaces en1 are tagged for VLAN 10 by the VLAN Ethernet adapter en1 in AIX.
- Only packets for VLAN 10 are received by the network interfaces en1.
- Packets sent through en0 are not tagged by AIX, but are automatically tagged for the VLAN specified as PVID by the POWER Hypervisor.
Only packets for the VLAN specified as PVID are received by the network interfaces en0.

In the configuration shown in Figure 2-32 on page 99, the Virtual I/O Server (VIOS) bridges both VLAN 1 and VLAN 10 through the Shared Ethernet Adapter (SEA) to the external Ethernet switch. But the Virtual I/O Server itself can only communicate with VLAN 1 through its network interface en2 attached to the SEA. Because this is associated with the PVID, VLAN tags are automatically added and removed by the POWER Hypervisor when sending and receiving packets to other internal partitions through interface en2.

Table 2-6 summarizes which partitions in the virtual Ethernet configuration from Figure 2-32 on page 99 can communicate with each other internally through which network interfaces.

Table 2-6  Inter-partition VLAN communication

<table>
<thead>
<tr>
<th>Internal VLAN</th>
<th>Partition / network interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Partition 1 / en0</td>
</tr>
<tr>
<td></td>
<td>Partition 2 / en0</td>
</tr>
<tr>
<td></td>
<td>Virtual I/O Server / en2</td>
</tr>
<tr>
<td>2</td>
<td>Partition 3 / en0</td>
</tr>
<tr>
<td></td>
<td>Partition 4 / en0</td>
</tr>
<tr>
<td>10</td>
<td>Partition 1 / en1</td>
</tr>
<tr>
<td></td>
<td>Partition 3 / en1</td>
</tr>
</tbody>
</table>

If the Virtual I/O Server is required to communicate with VLAN 10 as well, then it would need to have an additional Ethernet adapter and network interface with an IP address for VLAN 10, as shown on the left of Figure 2-33 on page 101. A VLAN-unaware virtual Ethernet adapter with a PVID only, as shown in the left of Figure 2-33 on page 101, would be sufficient; there is no need for a VLAN-aware Ethernet adapter (ent4), as shown in the center of Figure 2-33 on page 101. The simpler configuration with a PVID only would be effective, since the Virtual I/O/Server already has access to VLAN 1 through the network interface (en2) attached to the SEA (ent2). Alternatively, you could associate an additional VLAN Ethernet adapter (ent3) to the SEA (ent2), as shown on the right in Figure 2-33 on page 101.
Communication with external networks

The Shared Ethernet Adapter (SEA) of Figure 2-32 on page 99 is configured with default PVID 1 and default adapter ent1. This means that untagged packets or packets with VID=1 that are received by the SEA from the external network are forwarded to adapter ent1. The virtual Ethernet adapter ent1 has the additional VID 10. Thus, packets tagged with VID 10 will be forwarded to ent1 as well.

The handling of outgoing traffic to the external network depends on the VLAN tag of the outgoing packets:

- Packets tagged with VLAN 1, which matches the PVID of the virtual Ethernet Adapter ent1, are untagged by the POWER Hypervisor before they are received by ent1, bridged to ent0 by the SEA, and sent out to the external network.

**Note:** Although it is possible to configure multiple IP addresses on a Virtual I/O Server, it is recommended to have no more than one, because some commands of the command line interface make this assumption. Thus, a Virtual I/O Server should have one IP address or no IP address.

An IP address is necessary on a Virtual I/O Server to allow communication with the HMC through RMC, which is a prerequisite to perform dynamic LPAR operations. Thus, we recommend having exactly one IP address on a Virtual I/O Server, if you want to be able to use dynamic LPAR with the Virtual I/O Server.
Packets tagged with a VLAN other than the PVID 1 of the virtual Ethernet adapter ent1, such as VID 10, are processed with the VLAN tag unmodified.

In the virtual Ethernet and VLAN configuration example of Figure 2-32 on page 99, partition 1 and partition 2 have access to the external Ethernet through network interface ent0 using PVID 1.

- Since packets with VLAN 1 are using the PVID, the POWER Hypervisor will remove the VLAN tags before these packets are received by ent0 of partition 1 and 2.
- Since VLAN 1 is also the PVID of ent1 of the SEA in the Virtual I/O Server, these packets will be processed by the SEA without VLAN tags and will be send out untagged to the external network.
- Therefore, VLAN-unaware destination devices on the external network will be able to receive the packets as well.

Partition 1 and Partition 3 have access to the external Ethernet through network interface en1 and VLAN 10.
- These packets are sent out by the VLAN Ethernet adapter ent1, tagged with VLAN 10, through the physical Ethernet adapter ent0.
- The virtual Ethernet adapter ent1 of the SEA in the Virtual I/O Server also uses VID 10 and will receive the packet from the POWER Hypervisor with the VLAN tag unmodified. The packet will then be sent out through ent0 with the VLAN tag unmodified.
- Therefore, only VLAN-capable destination devices will be able to receive these.

Partition 4 has no access to the external Ethernet.

Table 2-7 summarizes which partitions in the virtual Ethernet configuration from Figure 2-32 on page 99 can communicate with which external VLANs through which network interface.

<table>
<thead>
<tr>
<th>External VLAN</th>
<th>Partition / network interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Partition 1 / en0</td>
</tr>
<tr>
<td></td>
<td>Partition 2 / en0</td>
</tr>
<tr>
<td></td>
<td>Virtual I/O Server / en2</td>
</tr>
<tr>
<td>10</td>
<td>Partition 1 / en1</td>
</tr>
<tr>
<td></td>
<td>Partition 3 / en1</td>
</tr>
</tbody>
</table>
If this configuration must be extended to enable Partition 4 to communicate with devices on the external network, but without making Partition 4 VLAN-aware, the following alternatives could be considered:

- An additional physical Ethernet adapter could be added to Partition 4.
- An additional virtual Ethernet adapter ent1 with PVID=1 could be added to Partition 4. Then Partition 4 would be able to communicate with devices on the external network using the default VLAN=1.
- An additional virtual Ethernet adapter ent1 with PVID=10 could be added to Partition 4. Then Partition 4 would be able to communicate with devices on the external network using VLAN=10.
- VLAN 2 could be added as additional VID to ent1 of the Virtual I/O Server partition, thus bridging VLAN 2 to the external Ethernet, just like VLAN 10. Then Partition 4 would be able to communicate with devices on the external network using VLAN=2. This would work only if VLAN 2 is also known to the external Ethernet and there are some devices on the external network in VLAN 2.
- Partition 3 could act as a router between VLAN 2 and VLAN 10 by enabling IP forwarding on Partition 3 and adding a default route via Partition 3 to Partition 4.

2.7.5 PowerVM Enterprise Edition and virtual Ethernet

A major additional capability of PowerVM Enterprise Edition is PowerVM Live Partition Mobility. PowerVM Live Partition Mobility allows you to migrate running AIX and Linux partitions (and their hosted applications) from one physical server to another without disrupting the infrastructure services. The migration operation maintains complete system transactional integrity. The migration transfers the entire system environment, including processor state, memory, attached virtual devices, and connected users.

For more information, refer to IBM System p Live Partition Mobility, SG24-7460.

**Note:** PowerVM Live Partition Mobility is *not* supported by the PowerVM Standard Edition feature.

PowerVM Live Partition Mobility supports two types of migration; *Inactive migration*, and *Active migration*.

- Inactive migration moves the definition of a powered off logical partition from one system to another along with its network and disk configuration.
- Active migration moves a running logical partition from a source system to a destination system with no disruption of partition operation or user service.