Introduction to Layer 3 Switching

In the previous chapter, you were introduced to the concept of inter-VLAN routing, which is required to enable hosts that belong to different VLANs on the same LAN network to communicate with each other. Implementing inter-VLAN routing introduces several benefits, which include the following:

- Reduces broadcast domains, increasing network performance and efficiency.
- Multilayer topologies based upon inter-VLAN routing are much more scalable and implement more efficient mechanisms for accommodating redundant paths in the network than equivalent flat Layer 2 topologies that rely on spanning tree alone.
- Allows for centralized security access control between each VLAN.
- Increases manageability by creating smaller "troubleshooting domains," where the effect of a faulty network interface card (NIC) is isolated to a specific VLAN rather than the entire network.

Of course, all of these features must be provided with a very important caveat—inter-VLAN routing should not affect performance, as users expect high performance from the LAN.

A popular approach to providing the benefits of inter-VLAN routing and also ensuring the performance of the LAN is not degraded has been to implement Layer 3 switches, which are essentially Layer 2 switches with a routing engine that is designed to specifically route traffic between VLANs in a LAN environment. Using Layer 3 switches for inter-VLAN routing as opposed to traditional routers is popular (and recommended) for the following reasons:

- **Performance versus Cost**—Layer 3 switches are much more cost effective than routers for delivering high-speed inter-VLAN routing. High performance routers are typically much more expensive than Layer 3 switches. For example, a Catalyst 3550-24 EMI switch sets you back $4,990 U.S. list, which provides a packet forwarding rate of 6.6 million packets per second with 24 * 10/100BASE-T ports and 2 * 1000BASE-X ports. A Cisco 7300 router with an NSE-100 engine provides a packet forwarding rate of 3.5 million packets per second, but sets you back $22,000 U.S. list and has only 2 * 1000BASE-T ports in its base configuration. Of course, the Cisco 7300 router has many more features and can support a wide variety of WAN media options; however, many of these extra features are not required for inter-VLAN routing.
- **Port density**—Layer 3 switches are enhanced Layer 2 switches and, hence, have the same high port densities that Layer 2 switches have. Routers on the other hand typically have a much lower port density.
- **Flexibility**—Layer 3 switches allow you to mix and match Layer 2 and Layer 3 switching, meaning you can configure a Layer 3 switch to operate as a normal Layer 2 switch, or enable Layer 3 switching as required.

Layer 3 switching is cheap because Layer 3 switches are targeted specifically for inter-VLAN routing, where only Ethernet access technologies are used in high densities. This makes it easy for Layer 3 switch vendors such as Cisco to develop high performance Layer 3 switches, as vendors can develop hardware chips (known as *application-specific integrated circuits* or *ASICs*) that specifically route traffic between Ethernet networks, without having to worry about the complexities of also supporting WAN technologies such as Frame Relay or ATM. Routing over WAN networks can still be supported, simply by plugging a traditional router that connects to the WAN networks into the LAN network. Figure 6-1 illustrates the concept of Layer 3 switching.

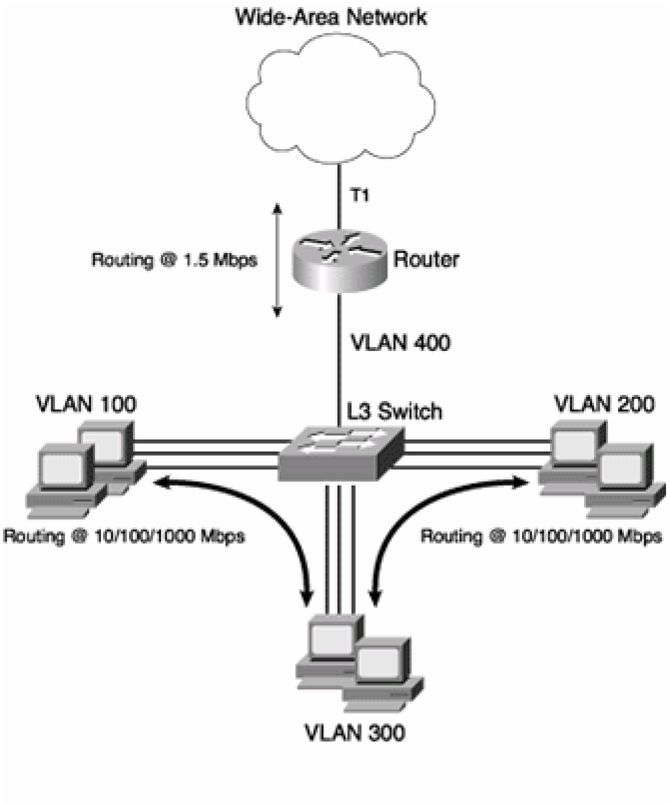


Figure 6-1 Layer 3 Switching

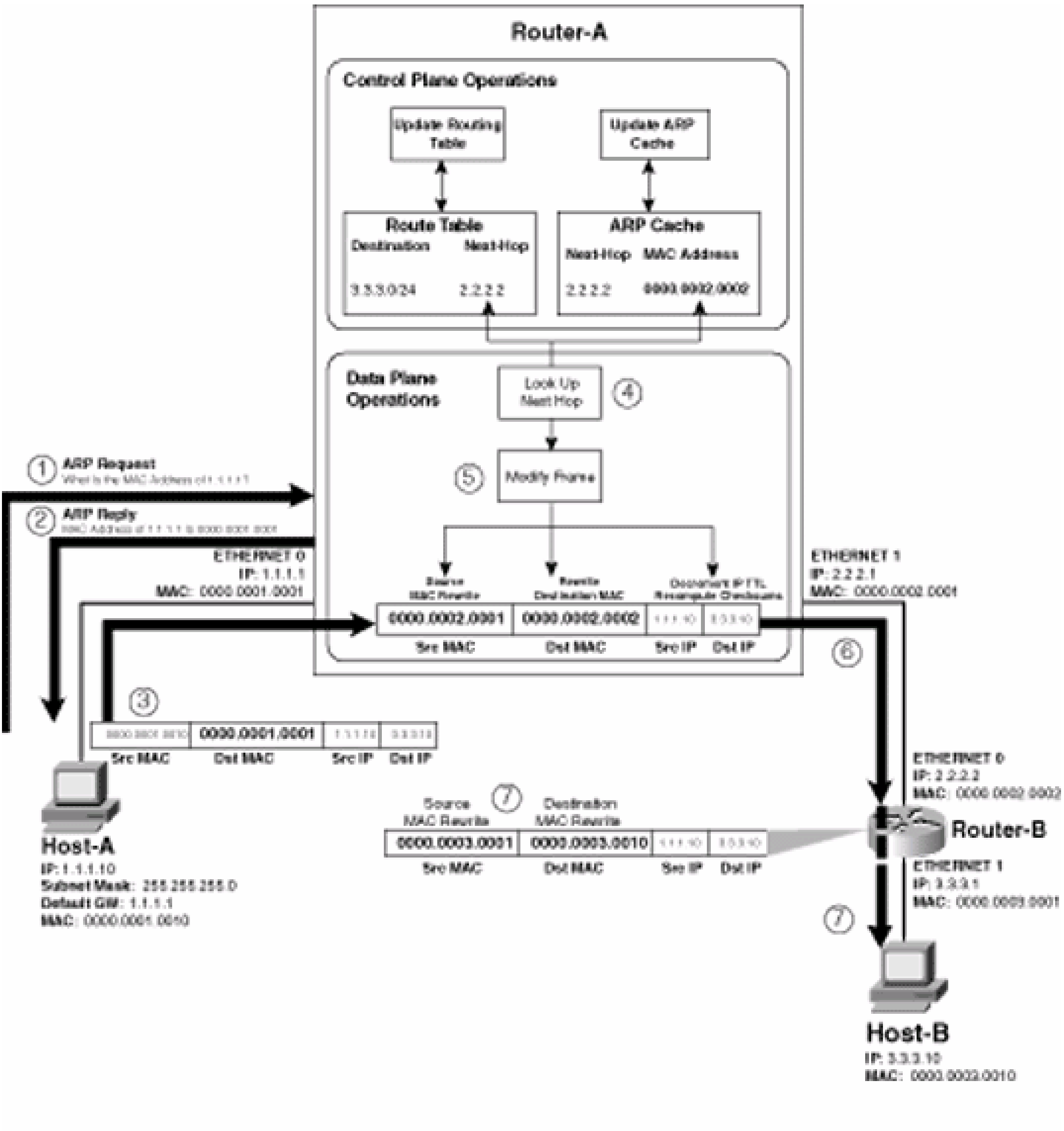
In Figure 6-1, a L3 switch provides switched LAN connections for each device in the network. Three user VLANs are present, and a routing engine on the L3 switch enables communications between each VLAN. The L3 switch possesses specialized hardware chips called application-specific integrated circuits (ASICs) that are preprogrammed and designed to route between Ethernet ports at high speed. A traditional router is connected to the L3 switch and handles the routing of any traffic that needs to be sent across the WAN. Because the L3 switch does not need the flexibility required of the router to support different WAN protocols, it can use ASICs to route traffic at the 100-Mbps speeds expected of the LAN network. The router in the network is designed to handle the requirements of routing at T1 (1.5 Mbps) speeds and would cause a bottleneck if it had to route between VLANs, as routing is performed in software, not hardware. Of course, you could purchase an expensive high-performance router with three Ethernet ports and a T1 interface; however, the cost associated with this approach is much higher. The cost associated with adding more routed Ethernet ports to the router (e.g., if a new VLAN was added to the network) is also high.

Layer 3 Routing Versus Layer 3 Switching

It is important to understand the difference between Layer 3 routing and Layer 3 switching. Both terms are open to some interpretation; however, the distinction between both can perhaps be best explained by examining how an IP packet is routed. The process of routing an IP packet can be divided into two distinct processes:

- **Control plane—**The control plane process is responsible for building and maintaining the IP routing table, which defines where an IP packet should be routed to based upon the destination address of the packet, which is defined in terms of a next hop IP address and the egress interface that the next hop is reachable from. Layer 3 routing generally refers to control plane operations.
- **Data plane—**The data plane process is responsible for actually routing an IP packet, based upon information learned by the control plane. Whereas the control plane defines where an IP packet should be routed to, the data plane defines exactly how an IP packet should be routed. This information includes the underlying Layer 2 addressing required for the IP packet so that it reaches the next hop destination, as well as other operations required on for IP routing, such as decrementing the time-to-live (TTL) field and recomputing the IP header checksum. Layer 3 switching generally refers to data plane operations.

Figure 6-2 illustrates the differences between control plane operation and data plane operation by providing an example of how an IP packet is routed.



.

Figure 6-2 Control Plane and Data Plane Operation

NOTE

Some Cisco Catalyst Layer 3 switches support the Layer 3 switching of Internetwork Packet Exchange (IPX) packets as well. For this chapter, the discussion focuses purely on IP packets In Figure 6-2, Host-A is sending an IP packet to Host-B over a LAN network that includes a couple of routers. The following describes the events that occur in Figure 6-2.

Step 1 Host-A (1.1.1.10) needs to send an IP packet to Host B (3.3.3.10). Host-A determines (by considering its own IP address, its subnet mask, and the IP address of Host-B) that Host-B is a non-local host and, therefore, must send the IP packet to the configured default gateway of 1.1.1.1 (Router-A). Because Host-A is connected to the network via Ethernet, Host-A must deliver the original IP packet in an Ethernet frame to Router-A. To place the packet in an Ethernet frame that can be delivered to Router-A, Host-A must know the MAC address of Router-A's 1.1.1.1 interface. Host-A checks the local Address Resolution Protocol (ARP) cache to see whether or not it knows the MAC address of Router-A (1.1.1.1). Assuming Host-A does not know the MAC address, Host-A broadcasts an ARP request, which is sent to all devices on the local LAN and asks for the MAC address associated with the IP address 1.1.1.1.

Step 2 Because Router-A is configured with an IP address of 1.1.1.1 on the interface attached to Host-A, it responds to the ARP request by sending a unicast ARP reply, which provides its MAC address (0000.0001.0001).

Step 3 Host-A can now encapsulate the IP packet in an Ethernet frame and send it to Router-A. The destination MAC address of the frame is the MAC address of Router-A, which ensures that Router-A receives the IP packet contained within for routing. The destination IP address, however, is not that of Router-A; it's that of Host-B, the true eventual destination of the packet (in other words, the IP addresses in the packet are not modified).

Step 4 Router-A receives the Ethernet frame and the data plane operations begin. For Router-A to forward the packet on to the appropriate next hop, it must know who the next hop is and the MAC address of the next hop. To determine the next hop, the router inspects the destination IP address of the IP packet (IP routing is always based upon the destination IP address). Router-A references the local route table for an entry that matches the destination IP address (3.3.3.10) and finds that 3.3.3.0/24 is reachable via a next hop IP address of 2.2.2.2 (Router-B).

Step 5 Because Router-A is connected to Router-B via Ethernet, Router-A must send the IP packet inside an Ethernet frame addressed to Router-B. To determine the MAC address associated with the next hop router, the local ARP cache on the router is checked to see if an entry exists for the IP address of the next hop. If no entry exists, then the router must generate an ARP request, asking for the MAC address associated with the next hop IP address (this is a control plane operation). Once the correct destination MAC address is known, the routed frame destination MAC address can be rewritten. The source MAC address is also rewritten to the MAC address of the Ethernet 1 interface on Router-A so that Router-B knows it received the frame from Router-A. It is this process of rewriting the frame MAC addresses that represents the key concept of data plane operations—A router does not modify the source or destination IP addresses of IP packets that are being delivered, but rather it must *rewrite* the destination and source MAC address so that the IP packet can be delivered over the LAN to the next hop.

NOTE

Router-A actually does have to modify some information in the IP header. Router-A must decrement the IP time-to-live (TTL) field and also must recompute the IP header checksum, since the TTL field has been changed. IP addressing might also be modified if network address translation (NAT) is configured; however, this operation is performed by a separate process outside of the control plane and data plane operations of routing.

Step 6 The rewritten Ethernet frame containing the IP packet is sent to Router-B.

Step 7 Router-B receives the frame from Router-A and examines the destination IP address of the packet. Because the destination IP address is that of a host that is locally connected, Router-B can complete the delivery by sending the packet to Host-B. Because Host-B is connected via Ethernet to Router-B, Router-B must send the IP packet inside an Ethernet frame addressed to Host-B. The same rewrite of the destination (and source) MAC address that was described in Step 5 takes place, and the frame is delivered to its final destination, Host-B.

NOTE

It is important to understand that the MAC addresses are specific only to each local LAN. For example, Host-A does not know and does not need to know Host-B's MAC address or even Router-B's MAC address. Host-A needs to know only the MAC address of Router-A so that it can deliver IP packets in Ethernet frames locally to Router-A, with Router-A then forwarding the packet on appropriately and with this process occurring on a hop-by-hop basis until the final destination is reached.

Control Plane and Data Plane Implementation

Control plane operations require an understanding of routing protocols and hence require some intelligence that is capable of supporting the complex algorithms and data structures associated with protocols such as Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP). Depending on the routing protocol(s) configured, the control plane operations required might vary dramatically between different routing devices. On the other hand, data plane operations are simple and fixed in their implementation because how a packet is routed is the same, regardless of the routing protocol that was used to learn where a packet should be routed. Although data plane operations are simple, they are also performed much more frequently than control plane operations because data plane operations must be performed for every packet that is routed, while control plane operations must be performed only for routing topology changes once the routing table is built. This means that the performance of the data plane implementation ultimately dictates how fast a routing device can route packets.

Because control plane operations are complex, most vendors use a general purpose CPU capable of supporting a high-level programming language so that vendors can easily develop and maintain the complex code associated with support the various routing protocols. In this respect, the control plane is implemented in *software,* which means that code (software) developed from a high-level programming language provides control plane operation. Both traditional routers and Layer 3 switches normally take the same approach to implementing the control plane operations associated with IP routing, using software that requires a general purpose CPU.

In contrast to control plane operations, data plane operations are very simple. In fact, the data plane operations required can be presented in a single table. Table 6-1 describes the data plane operations that must take place, assuming a packet is addressed from a host called Host-A to another host called Host-B and is sent via a router.

Table 6-1 Data Plane Operations Required on Received Frames

	Layer 2 Ethernet Header		**Layer 3 IP Header**				**Data**	**FCS**
	Destination MAC	Source MAC	Destination IP	Source IP	TTL	Checksum		
Received Frame	Router MAC Address	Host-A MAC Address	Host-B	Host-A	n	value1		
Rewritten Frame	Next Hop MAC Address	Router MAC Address	Host-B	Host-A	n-1	value2		

In Table 6-1, the details of the received frame are indicated and then the details required for the rewritten frame that is transmitted after routing are shown. Notice that the following fields must be modified for the rewritten frame that is forwarded to the next hop routing device:

- **Destination MAC address—**The MAC address of the next hop must be written to the rewritten frame.
- **Source MAC address—**The source MAC address must be written to the MAC address of the router.
- **IP TTL—**This must be decremented by one, as per the normal rules of IP routing.
- **IP Header Checksum—**This must be recalculated, as the TTL field changes.

The process of how the data plane operations shown in Table 6-1 are implemented is where the difference between a traditional router and Layer 3 switch lie. A traditional router uses the same general purpose CPU used to perform control plane operations to also implement data plane operations, meaning data plane operations are handled in software. A Layer 3 switch on the other hand uses an ASIC to perform data plane operations because it is very easy to program the very simple operations required for the data plane into an ASIC. In this respect, the data plane is implemented in hardware because a series of hardware operations are programmed into the ASIC that perform the data plane operations required for routing a packet.

NOTE

It should be noted that many high-end routers use ASICs for data plane operations in a similar fashion to Layer 3 switches. In fact, much of the ASIC technology used in Layer 3 switches is derived from the ASICs used in high-end routers.

So how does this affect performance? Well, a general purpose CPU is designed to support many different functions, where as an ASIC is designed to support a single function or a handful of specific functions such as performing the data plane operations required to route a packet. This means that an ASIC can operate much faster because the internal architecture of the ASIC can be optimized just to perform the operations required for data plane operations, whereas a general purpose CPU must be designed to support a series of generic functions that do not relate to data plane operations whatsoever (as the CPU must support other applications). A high-level language combines the generic functions of the general purpose CPU to provide the higher specific functions required to perform data plane operations. This approach allows flexibility but comes at the price of performance. Hence, a Layer 3 switch that performs data plane operations using ASICs route packets much faster than a traditional router that performs data plane operations using a general purpose CPU.

NOTE

The term *software* when applied to Layer 3 routing means that a general purpose CPU performs routing, along with other tasks such as system maintenance and providing command-line access. The term *hardware* when applied to Layer 3 switching means an ASIC dedicated to the process of Layer 3 switching, whose sole purpose in life is to route packets.

Hardware-Based Layer 3 Switching Architectures

Although the data plane operations required for routing IP packets can easily be accelerated by the use of ASICs, it is important to understand that a fundamental requirement for data plane operation is the process of determining the next hop IP address for the destination IP address of the packet and the MAC address associated with the next hop so that the correct destination MAC address can be written to the rewritten frame. The components that implement data plane operations must "look up" this information (see the lookup operation in Figure 6-2); this lookup operation in itself can become a bottleneck. To ensure the lookup process does not significantly delay the rewrite processes of data plane operation, Layer 3 switches use specialized data structures that allow for fast lookups. These data structures can be split into two categories:

- **Route cache—**A route cache is populated with information that defines how to Layer 3 switch frames associated with a particular *flow.* A flow uniquely identifies specific traffic conversations in the network (e.g., one flow might be Host-A communicating with Host-B, while another flow might be Host-A communicating with Host-C), and each flow entry contains the required information to Layer 3 switch packets received for that flow. The flow entries are built by routing the first packet in software, with the relevant values in the rewritten first frame used to fill out the required information for a flow entry. Subsequent packets associated with the flow are then Layer 3 switched in hardware based upon the information learned in the flow entry. Cisco's implementation of route caching on Cisco Catalyst switches is called *Multilayer switching* (*MLS*), and is discussed in more detail in Scenario 6-1.
- **Optimized route lookup table—**One approach to the lookup process could be to use the routing table; however, this contains information not relevant to data plane operations, such as the routing protocol that learned a route, metric associated with a route, and the administrative distance of a route. The routing table also does not contain MAC address information for the next hop. This must be determined either via a control plane operation (using ARP) or by reading the ARP cache. Next-generation

> Cisco Catalyst Layer 3 switches use an optimized route lookup table, which organizes only the required routing information for data plane operations (e.g., destination prefix, next hop, egress interface) and also includes a pointer to another optimized adjacency table, which describes the MAC address associated with the various next hop devices in the network. Cisco's implementation of using optimized route lookup tables on Cisco Catalyst switches is called *Cisco Express Forwarding* (*CEF*) and is discussed in more detail in Scenario 6-2 and Scenario 6-4.

It is important to note that in addition to possessing a high performance lookup mechanism, many Layer 3 switches also possess specialized hardware that can be used to provide QoS classification and security access control (using access control lists) for packets at the same time the next hop lookup is being implemented. This means that these features can be turned on with affecting performance.