CHAPTER 5

VLANs and Layer 3 Switching

This chapter covers the relatively new topics of virtual LANs (VLANs) and Layer 3 switching. The first part of this chapter discusses the concepts and benefits of VLANs and the different VLAN implementation methods, such as port, protocol, MAC address, and IP Subnet. We look at distributed VLANs, which require some kind of trunking. We look at the proprietary VLAN trunking method from Cisco called Interswitch Link (ISL), as well as the new IEEE 802.1Q VLAN tagging standard. We then look at the different methods of connecting VLANs.

We also examine the IEEE 802.1p priority switching technology that was designed to improve the delivery of time-critical data. Whereas 802.1p is not directly related to VLANs, the 802.1p standard utilizes the additional VLAN tagging field to provide for improved delivery of time-critical data. Therefore, this chapter is a good place to deal with it.

The second part of this chapter discusses Layer 3 switches. We first cover some of the basics of routing; then we attempt to explain the current hype surrounding Layer 3 switches, which are merely hardware-based IP or IPX LAN routers. We contrast the different types of Layer 3 switches, namely packet-to-packet, and a host of other methods that rely on cut-through methods. (For more information on cut-through methods, see Chapter 4, "Layer 2 Ethernet Switching.")

Next, we take an in-depth look at Layer 3 switches. Layer 3 switches are replacing classic IP LAN routers in certain places, because they offer more performance at significantly lower prices with easier configuration. Different VLANs need to be connected via a Layer 3 device, such as a classic router. The

required use of routers has slowed the adoption rate of VLANs tremendously. The recent emergence of fast, easy to configure, and affordable Layer 3 hardware-based switches has made VLANs a popular mainstream tool for LAN managers everywhere.

Finally, we discuss Layer 4 switching, which is the latest hot concept. We also discuss why Layer 4 switching isn't really switching at all.

Before we begin, we need to point out a few things:

- We cover a lot of information in this chapter. To discuss both VLANs and Layer 3 switches in detail in one chapter is ambitious. Therefore, this chapter doesn't go into too much detail. Entire books have been written on just VLANs or Layer 3 switching. Many excellent white papers from the vendor community also are available on the Web. Please refer to Appendixes B and C for suggestions for further reading.
- VLANs are not specific to Ethernet but are based on the OSI Layer 2. As such, you can apply VLANs to any frame-based LAN technology that follows the OSI model. For example, VLANs are already a feature on some of today's FDDI or Token Ring switches. The standards we discuss in this chapter, such as 802.1Q/p, are not 802.3/Ethernet specific; they were designed to work for Token Ring and other frame-based LAN standards as well.
- We discuss Layer 3 aspects of the OSI model. Note that, like VLANs, Layer 3 switching technology could be built into any OSI-compliant LAN hardware standard. Networking vendors chose to focus their Layer 3 switching efforts on Ethernet, and Fast and Gigabit Ethernet in particular. For example, we quite possibly will see some Layer 3 Token Ring switches one day.
- TCP/IP is the most popular Layer 3 networking protocol today and is fast becoming the de facto standard. We assume you are somewhat familiar with the routing aspects of TCP/IP. We included a number of references for additional reading.

Before we delve into VLANs in more detail, let's take a small detour into the frame world.

Unicast, Multicast, and Broadcast Frames

This is a quick tutorial on the subject of unicast, multicast, and broadcast frames. These three frame transmissions differ in terms of their destination address.

Unicast Frames

Ethernet frames typically are sent from a particular source address (SA) to a specific destination address (DA), which is a one-to-one transmission. This is known as *unicast*, or *unique address broadcast*. The DA field in a unicast frame is the MAC address of the destination and is always a unique 6-byte number. An example of a unicast DA is 00-A0-EF-12-34-56 (hex). A switch directly forwards a unicast frame from source port to destination port. Most of today's LAN traffic consists of point-to-point transmissions.

Multicast Frames

Sometimes, the same source data needs to be sent to multiple receivers. This could be an email sent to everyone in a specific department within a company or a company-wide mailing list. Newer applications, such as server-based audio or video streaming (covered in Chapter 7, "Bandwidth: How Much Is Enough?"), also fall into this category of one-to-many communication. From a bandwidth-usage perspective, broadcasting to multiple users at once is much more efficient than generating multiple individual unicast transmissions for every individual user. The one-to-many transmission, broadcasting to multiple users at once, is called *multicasting*. One-to-many transmissions are becoming increasingly popular as new applications, such as server-based streaming, groupware, videoconferencing, and IP multicasting, become more mainstream. The receiver typically decides whether to join a specific multicast transmission. Multicasts are identified by means of a particular DA range of addresses. An example of a multicast DA is 01-80-C2-00-00-00 (which is a spanning tree multicast address).

Note

Many recent one-to-many applications utilize IP multicasting, which is different from Ethernet multicasting. We discuss IP multicasts later in this chapter.

When Layer 2 switches do not know the destination address of a particular frame, they flood or forward the frame to all ports. Although this is not a multicast, it is similar in nature because it creates a lot of extra traffic on all segments.

Broadcast Frames

Broadcasts are one-to-all transmissions addressed to everyone on the network. A variety of sources can generate broadcast frames:

- The spanning tree Bridge Protocol Data Unit (BPDU) (discussed in Chapter 4) is a typical example of a broadcast frame. This frame needs to be conveyed to all bridges and switches on the LAN to build one tree.

- Some network operating system (NOS) clients and servers, such as NetWare, use broadcasts to advertise their presence on the LAN. Most networking protocols, such as IPX, AppleTalk, and NetBIOS, make regular use of broadcasts to discover addresses, routers, and servers.
- IP routing protocols, such as RIP, SNMP, DVMRP, and ARP, use broadcasts extensively to discover routing paths, to exchange information about the optimal route, or to match IP and MAC addresses.

With broadcasts, the DA frame field is set to all *1s*. This indicates to all stations on the LAN that this particular frame is destined for everyone. All broadcast frames have the destination address field set to all 1s (or FF-FF-FF-FF-FF-FF, in hexadecimal notation).

Broadcast Storms

Like Ethernet collisions, broadcasts have received a bad reputation over the years because they often do not represent an actual data transmission but more network Layer 2 or 3 overhead.

A unicast transmission will occupy only the direct path from source to destination, whereas a broadcast will permeate all corners of a network. Thus, many people view broadcasts as wasted bandwidth. The truth is that broadcast transmissions cannot be eliminated: They are part of the normal workings of any network. Only excessive broadcast rates are a problem, and they are typically the result of too large a network or faulty hardware. Excessive broadcast rates affect the net bandwidth available to users, causing the network to become very sluggish.

Faulty hardware or an incorrectly configured network can lead to the network being overwhelmed with broadcasts, also known as a *broadcast storm*. A broadcast storm describes a situation in which the entire network is used to transmit broadcasts, leaving no bandwidth for regular traffic. This will result in timeouts and network errors, which is the equivalent of freeway "gridlock."

Really, only two possible things can create a broadcast storm:

- Bridge loops (such as that shown in Figure 4.5 in Chapter 4) running without the spanning tree algorithm (STA) in operation. This is very rare, because most new bridges and all switches include STA. (Alternatively, STA could be disabled on a bridge or switch.)
- The other, more likely, source of broadcasts is faulty hardware (NICs or switches). The device in question is malfunctioning and possibly sending out broadcast frames permanently.

Both scenarios are very unlikely these days; NICs now have an MTBF (mean time between failure) rate of well over 10 years, and all bridge switches include the STA.

Broadcast Domains

In a shared or repeated Ethernet segment, all nodes are in the same collision domain. A repeater is invisible to all nodes in a shared-media segment. Nodes operating in a shared environment collide with each other if they attempt a transmission at the same time. When two nodes collide, all other nodes hear the collision: hence the term *collision domain*. Collisions do not traverse a bridge or a switch; therefore, these devices form the edges or borders of a collision domain.

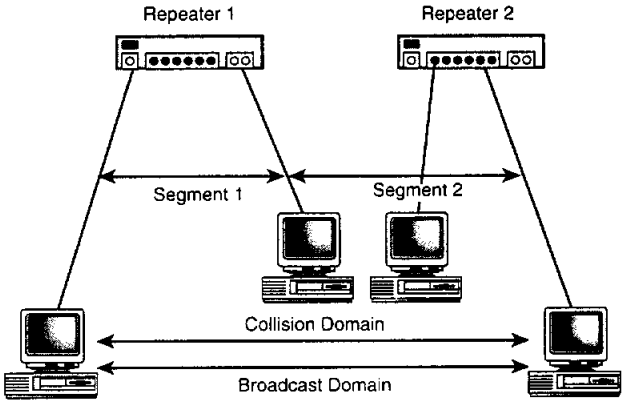
The broadcast domain is the network area that a broadcast frame will fill. As discussed in Chapter 4, switches or bridges operate at Layer 2 and blindly forward all broadcast traffic received (as well as all unknown destination address frames), making switches invisible to broadcasts. Figure 5.1 illustrates this.

Problems with Very Large Broadcast Domains

Theoretically, you can build very large networks with hundreds or even thousands of nodes using only Layer 2 switches. The term *flat network* is used because, from a hierarchical perspective, all these switches are on the same level: no higher level Layer 3 routers are present. In practice, the size of a flat, switched Layer 2 network has a limit. The following factors limit the size of a Layer 2 network:

- *Broadcasts grow with network size*—As previously discussed, broadcasts tend to travel everywhere within a switched network. Many NOSs and their associated protocols, such as NetWare, AppleTalk, LAN Manager, and LANServer, are rather chatty: They create a fair amount of broadcast traffic. That's because these network operating systems' protocols were designed to operate on local area networks where bandwidth has traditionally not been a problem. When enlarging a Layer 2 switched network from say 100 to 1000 users, the broadcast rate will grow at least tenfold, too. If your broadcast rate in a 100-user network is 2%, the 1000-user network will have a broadcast rate of at least 20%, which is a rather significant number. Even 20% broadcast traffic is tolerable, because it leaves you with almost 80% usable bandwidth in a switched environment. The good news is that TCP/IP is a very quiet (or, non-chatty) protocol because it was designed to operate over WAN links, where bandwidth is scarce. Therefore, as more networks migrate from other protocols to TCP/IP, the broadcast rate will decrease.

- *Control*—The big issue with large networks is that a broadcast storm will crash the entire network. Although broadcast storms are extremely rare, the occurrence of one will bring your entire network to its knees, and 1000 angry users is nothing to sneeze at.
- *Lack of IP addresses*—Networks using IP must contend with another issue: lack of IP addresses. In an IP environment, every node receives an IP address that is either permanently (statically) assigned or dynamically assigned via the Dynamic Host Configuration Protocol (DHCP). Only a maximum of 254 IP addresses can be assigned on a particular IP subnet, imposing an artificial limit on the maximum number of users in a particular broadcast domain.



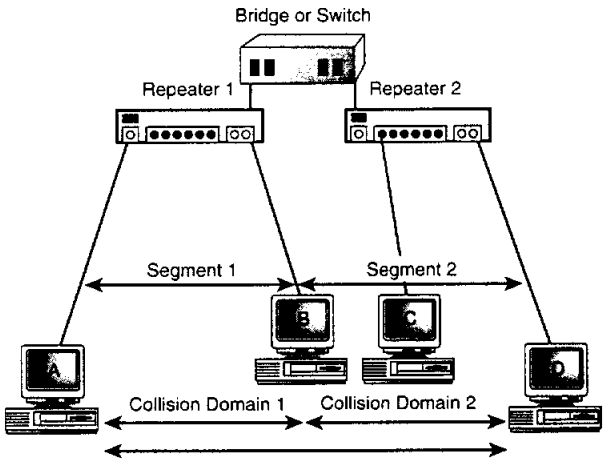


FIGURE 5.1 *For a repeated network, the collision domain and broadcast domain are the same. Introducing a bridge or switch creates two separate collision domains, yet all nodes still share the same broadcast domain.*

The Old Way of Dealing with Large Broadcast Domains: Routers

Historically, you connected LANs together over short distances with bridges. You used routers to connect LANs together over extended distances. You joined these different LANs via multiple routers, sometimes with different path choices. Routers set up an optimum routing path depending on various criteria, therefore the term *routers*. Routers operate at Layer 3 and don't forward broadcasts or multicasts automatically, so LAN managers started using routers to link similar LANs over shorter distances. As corporate networks became larger, routers started moving to the center of the network to segment larger, flat Layer 2 networks into smaller broadcast domains or subnets. Routers became the centerpieces for large enterprise networks, using architectures such as distributed or collapsed backbones (more on this in the second half of this book) to generate multiple smaller subnets.

This solves all of the issues of large flat Layer 2 networks: Every subnet can have 254 IP addresses, broadcasts are contained within a subnet, possible broadcast storms only affect one subnet, and protocol routing still occurs. (This assumes a particular type of IP address, namely a Class C address.) Figure 5.2 shows a switched network that has been subdivided by means of a router. The nodes A, B, C, and D (shown in Figure 5.2) physically reside in the same broadcast domain, and nodes E, F, and G reside in broadcast domain 2. The limits for broadcast domains 1 and 2 are the outer physical boundaries set by switch 1 and switch 2. If, for example, node D wanted to join broadcast domain 1, the node would have to be physically moved and connected to switch 1.

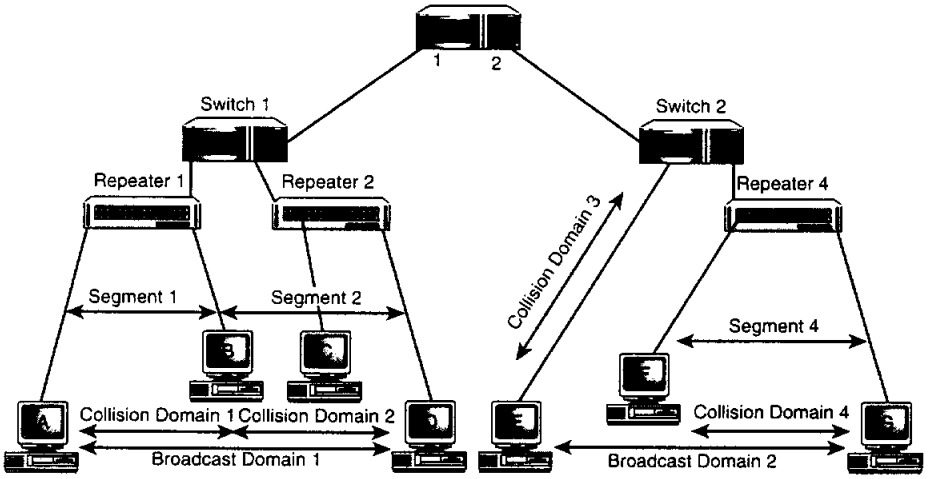


FIGURE 5.2 *Introducing a router into our previously shown network creates two different broadcast domains. Nodes A, B, C, and D are part of broadcast domain 1, whereas nodes E, F, and G are part of broadcast domain 2.*

Note

Interestingly, the broadcast rate in most pure Layer 2 networks is not excessive. What LAN managers fear most is the entire network coming to a halt. Unfortunately, broadcast storms can cause that, extremely rare as they are. So it is often the fear of broadcast storms, not of existing broadcast rates, that causes network managers to look for ways of dividing up larger Layer 2 networks into smaller broadcast domains.

So before you go out and buy more routers to segment your flat Layer 2 network to reduce broadcasts, we urge you to take a closer look at what exactly your broadcast traffic rate is. Chapter 7, "Bandwidth: How Much Is Enough?" discusses some tools that allow you to measure broadcasts. In our opinion, a 10% broadcast rate is quite acceptable.

Note

Routers often subdivide a larger flat network into multiple broadcast domains. The router is placed at the center of the network to accomplish this. There's another reason why routers often are found at the center of large networks. Historically, different operating systems have all used different networking protocols. Only recently has TCP/IP emerged as the clear leader. For many years, Novell NetWare ran on IPX, Digital used LAT, Microsoft favored NetBEUI or NetBIOS, and the UNIX world was synonymous with TCP/IP. Then, of course, Banyan, IBM, Apple, and the like all used different protocols again. So most corporate networks required a powerful protocol router somewhere so that everyone could communicate with everyone else. Putting this router at the edge of the network wasn't very efficient because it meant the routed traffic needed to traverse multiple switches twice. Thus, routers became the centerpieces of larger mixed-protocol environments.

VLANs

No textbook definition or standard describes a VLAN perfectly, but over the past few years, a commonly accepted definition has emerged. A VLAN is a logical grouping of nodes, consisting of clients and servers that reside in a common broadcast domain, without any router hops. The word *virtual* means that the LAN manager has artificially created the broadcast domain within the switching fabric. Whereas a VLAN is a single broadcast domain, the different nodes within this VLAN need not be physically connected to the same switch or even be in the same physical area. Nodes that are members of the same VLAN appear as though they are connected to one Layer 2 bridge or switch, sharing broadcasts, but they may not actually be. These different nodes may, in fact, be connected to different switches in different buildings altogether and connected only via a VLAN. Alternatively, these VLAN members may be connected to the same switch but may not be able to see all the broadcast traffic originating from other ports on that very same switch.

Figure 5.3 shows the previous network example from Figure 5.2 with two different VLANs.

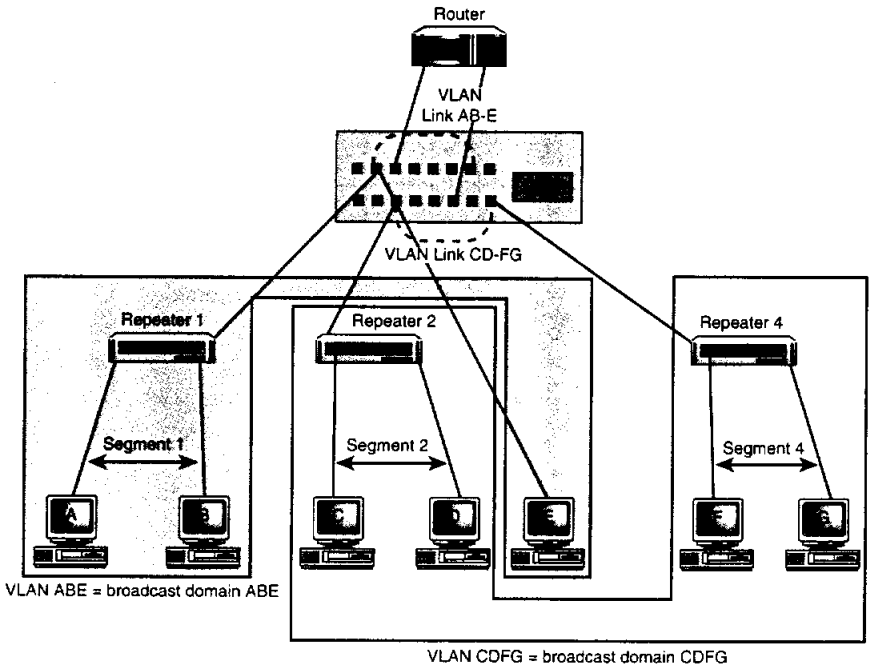


FIGURE 5.3 *Our two Layer 2 switches have been replaced with a single VLAN-capable switch. The VLAN-capable switch sets up the broadcast domains internally. Nodes A, B, and E now reside in VLAN ABE, and nodes C, D, F, and G share broadcast domain CDFG. A router will be required to communicate between VLANs ABE and CDFG.*

Benefits of VLANs

Let's look at some of the benefits of VLANs:

- *More bandwidth*—VLANs isolate broadcasts. Fewer broadcasts mean more bandwidth is available for regular unicast traffic.
- *Frees up your network from physical limitations*—A VLAN is a grouping of network nodes that share similar resources. Unlike a pure Layer 2–switched LAN, these resources do not have to be physically located next to each other in the network. For instance, a marketing department that is spread over Buildings HQ and SALES can all access the marketing department's shared servers, despite the fact that the server is located in the basement of the IT building. The benefit of a VLAN is obvious: shared resources don't need to be in the same physical area. Figure 5.4 shows this example.

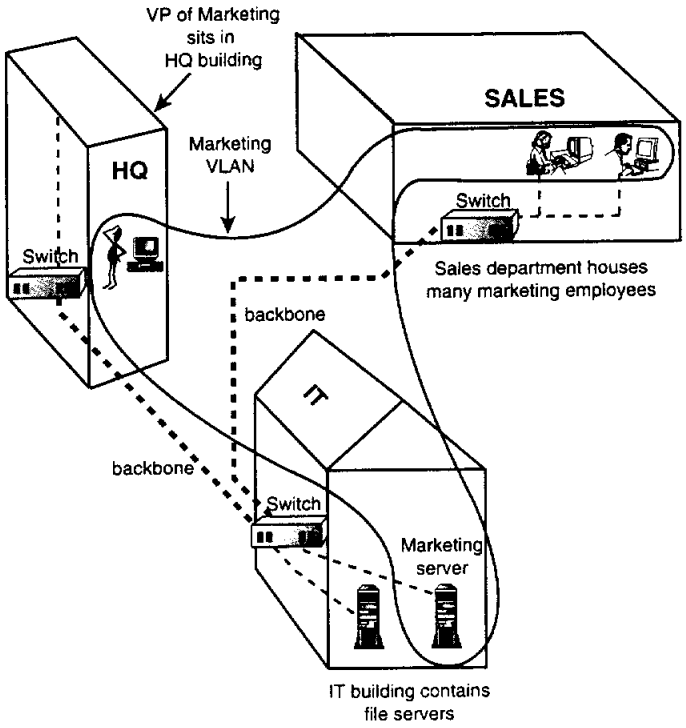


FIGURE 5.4 *A VLAN groups users and shared resources, often located in physically different areas. Shown is a marketing department spread out over two different buildings, accessing shared servers in a third building.*

- *Broadcast containment*—According to our definition, VLANs are synonymous with broadcast domains. Routers also contain broadcast traffic within a VLAN domain. Thus, VLAN-enabled switches can replace routers with the added benefit of being cheaper and easier to configure than routers, which saves you time and money.
- *Multicast containment*—Yes, VLANs will contain broadcasts. In addition, VLANs will also contain multicasts, reducing overall network traffic.
- *Easy changes*—Because VLANs enable the dynamic allocation of resources, VLANs also enable you to make changes easily. If a user wants to change office location from the HQ building to the IT building, this move can be easily accommodated by extending the VLAN to include the new office location in the IT building.

- *Easily shares resources*—A server or desktop can actually be part of multiple VLANs. This reduces the need to route traffic and provides greater flexibility and access to resources when and where they are needed.
- *Performance*—VLANs offer the capability of bandwidth on demand. If, for example, some users on a particular VLAN complain about the lack of bandwidth, you could create a new dedicated VLAN and move several users to this new VLAN, thereby improving performance. This new VLAN would involve less traffic, thereby improving performance.
- *Security*—In a flat switched environment, everyone shares broadcasts. If some of these broadcasts contain secure information, someone unauthorized could conceivably obtain access to confidential information. VLANs offer you the benefit of true broadcast security. If, for example, the HR department wants to keep all its traffic confidential, you could create a VLAN with just the HR staff as part of it. This benefit is questionable in our opinion, as very few broadcasts will contain confidential information. In general, higher level security measures are easier to set up and maintain.

VLAN Membership

So, how do you decide what nodes to group together in a particular VLAN? VLANs come in numerous different types, and choosing how to group nodes together is sometimes not easy. Going back to our definition of VLANs will give you some pointers on how to group nodes together: VLANs are broadcast domains set up to share resources. VLAN memberships are also sometimes referred to as *policies*.

How do these VLAN implementations help in the management of an enterprise network? We'll look at each briefly and give some pluses and minuses.

Port-Based VLANs

With port-based VLANs, you manually assign a switch port to a particular VLAN number. For example, you can assign switch port 8 to a VLAN called FINANCE. This typically means that another switch or repeater is attached to that port and all nodes on that port are in the same VLAN. This would make sense if all finance people work in close physical proximity to each other and all connect to the same repeater or switch hanging off port 8.

Alternatively, you can connect multiple VLAN switch ports to form a common VLAN. For example, switch port 1 could connect to the marketing people in building HQ, port 2 could connect to the marketing employees in the SALES building, and port 5 could connect to the servers in the IT building. Grouping these three ports into one VLAN sets up MARKETING VLAN. In this case, the

switch will analyze each received broadcast frame and forward it only to the ports that have been assigned the same VLAN. Unicast frames, however, will be forwarded only to the required port.

Figure 5.5 shows an example of a port-based VLAN.

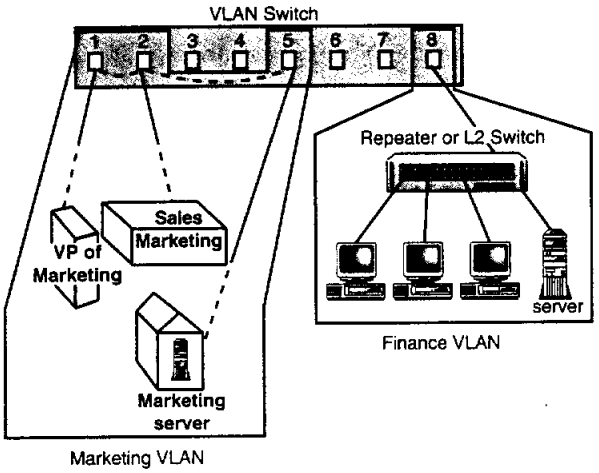


FIGURE 5.5 *Port-based VLANs are useful when the physical network resembles your logical layout.*

Advantages

Port-based VLAN setup is quick and easy to understand: There aren't that many ports on a switch.

Some first generation VLAN switches supported only this kind of configuration.

Disadvantages

You must manually keep track of all VLAN names, port numbers, and associated nodes connected.

A user physically connected to one port and moving to a different port will require reconfiguration.

MAC Address-Based VLANs

This method requires you to add individual MAC addresses manually to specific VLANs. This means that a given end-station, no matter where it is on a network, will be a member of that VLAN. The management of MAC-based VLANs is also manual: You need to track MAC Addresses and relevant VLAN

group numbers. For example, a VLAN called HR could consist of nodes with MAC addresses 00-A0-C9-12-3A-F3 and 00-60-06-44-35-D2.

Figure 5.6 shows an example of a MAC address-based VLAN.

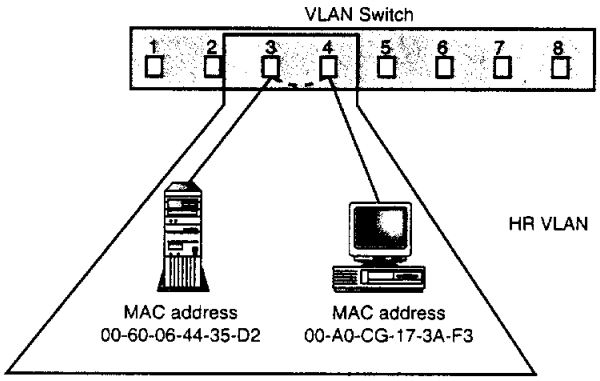


FIGURE 5.6 *VLANs are useful when the individual Ethernet MAC addresses are known and you have the time to enter them all manually.*

Advantages

Every NIC has a hardcoded MAC address. If you move the PC or notebook and the installed NIC, you automatically move the MAC address, too. The switch will retain the original VLAN membership, however, regardless of physical location. Therefore, you can use this VLAN management technique when your users connect laptops from anywhere in the building at any time.

Disadvantages

The biggest disadvantage is that every MAC address needs to be entered manually or added to a VLAN, which can be rather cumbersome.

While the idea of transparent moves sounds appealing, particularly for large notebook installations, there is a catch with docking stations, many of which have the NIC installed in them instead of in the notebooks. Thus, a user who moves to a different docking station will not be able to rejoin one's original VLAN group unless you account for this in the original VLAN setup.

If the NIC or the PC are faulty and is replaced, the switch VLAN configuration needs to be updated.

Layer 3 Protocol-Based VLANs

With Layer 3 protocol-based VLANs, you must be running more than one protocol. In this situation, you set up a VLAN based on what specific protocol is in

use. For example, VLAN 4 could be a grouping of all nodes using the IPX protocol, and VLAN 5 could comprise clients and servers using the IP protocol.

Figure 5.7 shows an example of a protocol-based VLAN.

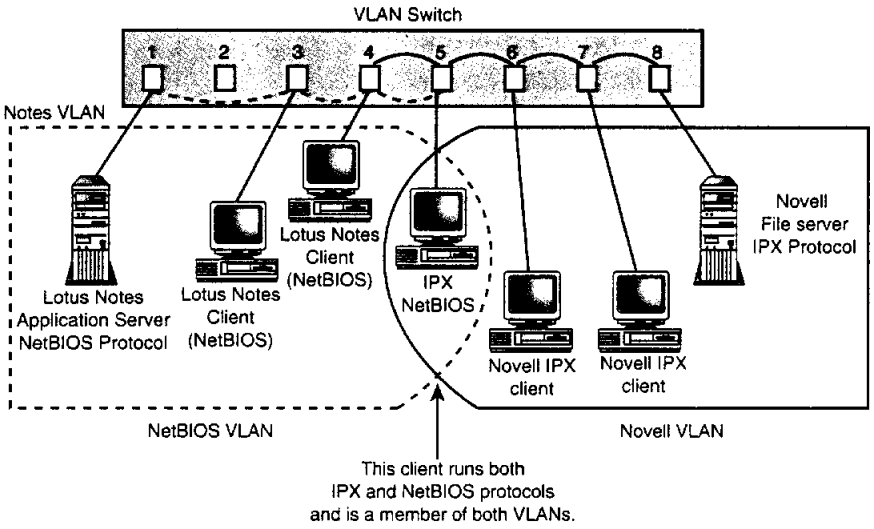


FIGURE 5.7 *Protocol-based VLANs are useful when different applications use different protocols.*

Advantage

Often, particular applications use a specific protocol. Segmenting traffic by protocol type allows you, in effect, to create an application-specific VLAN. Users can move to anywhere on the network and retain their VLAN membership so long as they keep using the same protocol.

Alternatively, you can segment by NOS server by choosing NetWare and NT as policies (via their respective protocols, such as IPX and IP). This is by far the most common use of this kind of VLAN.

Disadvantage

Analyzing the protocol type on every packet is very time-consuming. MAC- and port-based VLAN switching, on the other hand, involves almost no switch intervention at all and is very fast by comparison. Many networks are standardizing on IP everywhere, in effect creating one flat IP VLAN. That makes this VLAN method superfluous.

You can set up nonroutable protocol VLANs, but this doesn't make much sense because you cannot communicate outside of this VLAN. If you want to keep traffic local on a particular VLAN, though, this could work for you.

Layer 3/IP Network Address VLANs

IP is a protocol that assigns an individual address to every node: for example, 172.16.2.1 uniquely identifies a specific node. Different IP nodes can be grouped together to form one VLAN. This VLAN policy is similar to the MAC address scheme in that you need to know individual IP addresses. Whereas every node has a MAC address, not all nodes will run IP, and therefore, not all nodes will have an IP address. Because IP addresses often are assigned in ranges, setting up VLAN is preferred to coincide with an IP range of addresses, called a Subnet. Of course, this method of setting up VLANs won't work if you use DHCP because it assigns a different IP address every time a user connects to the network.

Layer 3/Network Subnet Address VLANs

Layer 3/network address VLAN assignment is similar to the protocol-based method in that it uses Layer 3 information to determine VLAN membership. This method works very well for IP LANs, where each individual node can have a unique IP subnet address. For example, a VLAN called Subnet 2 could comprise all nodes with the IP subnet address 172.16.2.0 and mask 255.255.254.0. This means that all 254 nodes within the IP address range 192.0.100.1–192.0.100.254 communicate within VLAN Subnet 2.

A VLAN switch might apparently act like a router when assigning VLAN membership according to Layer 3 information. This isn't the case. In fact, the VLAN switch acts as a grouper: It merely bundles a particular protocol or subnet traffic into a VLAN. In order to communicate among different protocols or subnets, a router is still required. That's why most, if not all, of today's VLAN switches are actually a combination of a VLAN switch and a Layer 3 switch.

Figure 5.8 shows an example of a subnet-based VLAN.

Advantages

This type of VLAN management works well if your VLAN grouping matches your physical IP subnet structure.

Subnet-based VLANs are probably the easiest to configure. In this case, a VLAN switch can often replace a classic router used for subnetting. This is the most backward-compatible version of VLANs with respect to replacing existing routers.

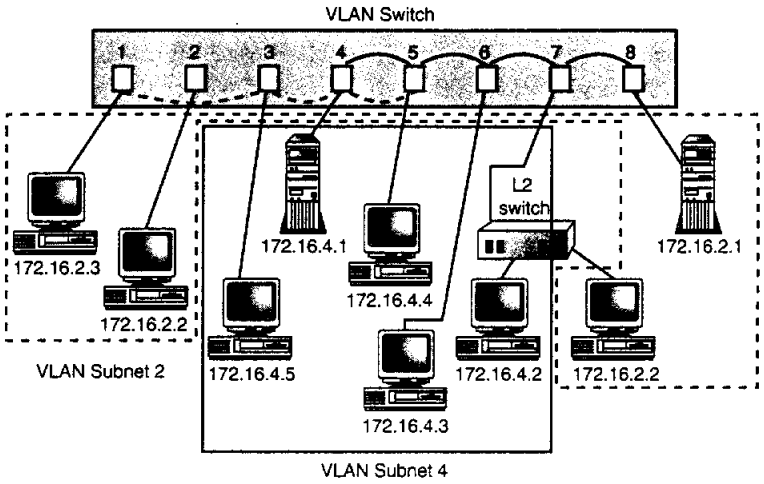


FIGURE 5.8 *Here is an example of a subnet-based VLAN. In this situation, a VLAN switch can often replace a classic router.*

Disadvantages

Network address-based VLANs only work for IP-based nodes.

Just as in MAC address-based VLANs, you need to keep track of individual IP addresses.

Multiple VLAN Membership

Some combinations of the previous VLAN policies are possible, too. Consider the following example: the HR VLAN is a MAC address-based VLAN (see Figure 5.6). Right now, switch ports 4 and 5 have only MAC node connected. You could add switch port 5 permanently to this VLAN. By assigning this port to the HR VLAN altogether, you can ensure that all nodes that are subsequently added to port 5 will also become part of the same VLAN. That way you don't have to worry about future configurations. If someone is connected to this port by accident, however, that person will have access to all the HR confidential data. This is basically a "mix-and-match" approach.

IP Multicast Address-Based VLANs

We can also identify a VLAN with an IP multicast address, which is a proxy address for a larger group of IP addresses. If a frame needs to go to this group of IP addresses, it is sent first to the proxy IP address and then forwarded to the entire group. Membership in the group is voluntary: Each desktop can determine if its IP address should be included in the group. This kind of VLAN

identification is useful in networks where video or audio data is being broadcast on the network and only a select few users are allowed or want to view or listen to the information.

An IP multicast VLAN is set up at Layer 3 or higher and has nothing to do with the actual VLAN switching hardware involved. IP multicast VLANs are also temporary, as a node can leave the multicast domain at any time. IP multicast VLANs don't provide broadcast containment: they are merely a grouping of nodes. By definition, IP multicast VLANs can also span WANs. Table 5.1 summarizes the different VLAN membership options. Figure 5.9 shows the VLAN configuration software of an actual Layer 3 switch. The switch shown has been configured with many different VLANs.

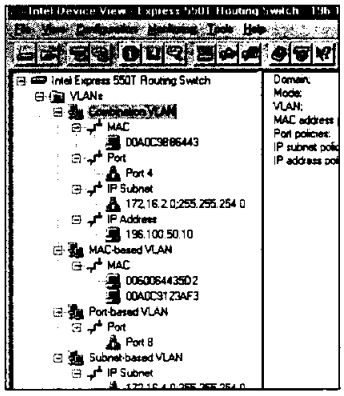


FIGURE 5.9 *VLAN configuration is made easy through good management software. For example, Intel's DeviceView switch management software contains an "Explorer" function that allows you to explore the switch by selecting individual VLANs. Double-clicking will reveal the different VLAN policies.*

TABLE 5.1 VLAN POLICY OR MEMBERSHIP OPTIONS

Type	Information Required	Example
Port-based	Switch port number	Port 8→Finance VLAN[1] Ports 1, 2, 5→Marketing VLAN
MAC-based	Client and server MAC	MAC addresses 00-A0-C9-12-3A-F3 addresses (from NIC) and 00-60-06-44-35-D2[→HR VLAN
Layer 3/protocol[2]	Protocol Type	NetBIOS packets→Notes VLAN IPX packets→Novell VLAN

continues

TABLE 5.1 CONTINUED

Type	Information Required	Example
Layer 3/IP subnet	IP Subnet and Mask	IP Subnet nodes with address 172.16.2.X and Mask 255.255.254.0 are on VLAN SUBNET2; nodes with 172.16.4.X and Mask 255.255.254.0 are on VLAN SUBNET4.
IP multicast-based	IP Multicast address	IP Multicast address X is and Mask associated with VLAN
Multiple VLAN membership	Combinations of above	Add port 5 to HR VLAN member ships (so far a MAC-based VLAN)

[1] *The → symbol indicates "assigned to."*
[2] *Nonroutable protocols, such as NetBIOS, are rarely used to define VLANs because you cannot communicate outside of this VLAN. You would more typically use IP and IPX to define protocol-based VLANs.*

The Evolution of VLANs and VLAN Membership

Let's take a look at the three generations of VLAN switches, from the first-generation, standalone switches to today's third-generation, distributed, standards-based VLAN switches.

First Generation: Standalone VLAN Switches

First-generation, VLAN-capable switches were limited to one switch only. An individual VLAN could encompass only a single switch, which severely limited VLAN deployment. For example, transparent user moves are one of the benefits of VLANs, but a user moving in a larger organization will unlikely be connected to the very same switch at the new location, which defeats the whole purpose of VLANs.

Second Generation: Distributed Proprietary VLANs

A VLAN spanning more than one switch is called a *distributed VLAN*. Second-generation VLAN switches could share VLAN membership information among different switches. In effect, a link or trunk between different switches or a server connection would contain traffic that was part of more than one VLAN. Cisco's Inter-Switch Link is probably the best-known distributed VLAN technology, having been licensed to many other vendors, such as Intel for its server NICs. ISL explicitly tags each frame with a particular VLAN group identifier. Other networking vendors have developed similar schemes. For example, 3Com switches use VLAN trunking (known as VLT), which is also a frame-tagging method like ISL. The issue with all these VLAN membership communication methods was that they were all proprietary: you could build a distributed VLAN only if using switches from a single vendor.

Figure 5.10 shows an example of a distributed VLAN.

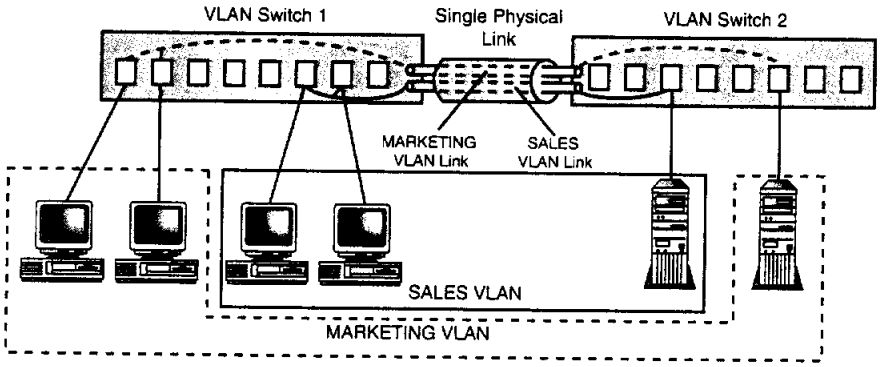


FIGURE 5.10 *The capability to distribute VLANs among different switches has really made VLANs a powerful concept for managing networks.*

Third Generation: Distributed, Standards-Based VLANs

The vendor community and the IEEE realized that a VLAN trunking standard was necessary because the proprietary nature of VLANs limited their acceptance. Two approaches were possible. First, different switches could exchange tables with relevant VLAN information, such as node MAC addresses, VLAN group membership, and so on. The spanning tree algorithm uses a similar table exchange method for broadcasting BPDU frames. This method's drawback is that it creates lots of extra LAN broadcast traffic, which VLANs are trying to limit. The other option seeks to identify each packet explicitly with a VLAN group membership number through a tagging approach.

In the end, the IEEE chose a standard that was based on Cisco's ISL frame-tagging developments. In the new 802.1Q standard, a special 4-byte tag is inserted into every Ethernet frame in between the source address and Length/Type field. This tag field contains two fields of information, the Tag Control Information (TIF) and Tag Protocol Identifier field (TPID):

- The Tag Control Information (TIF) field, which consists of the VLAN ID, the USER PRIORITY field, and a CFI bit.
- Twelve bits represent the VLAN ID, which determines the actual VLAN group. All switches in the network use this VLAN ID to communicate VLAN membership.
- The USER PRIORITY field is 3 bits long, and we will cover it in more detail in the next section.

- There is also the Canonical Format Indicator (CFI) bit, which is only used for Token Ring transmission. This bit indicates that the current frame is actually a Token Ring frame encapsulated in an Ethernet frame format.

The three fields of the TIF just mentioned add up to 16 bits, or 2 bytes. Then there is the Tag Protocol Identifier field (TPID), which is used for Token Ring, FDDI, and SNAP-encoded data transmissions. For Ethernet, it is of no concern and set to 81-00. The TPID field is also 16 bits, or 2 bytes, which brings the total for the total tag length to 32 bits, or 4 bytes.

Figure 5.11 shows the 802.1Q tag.

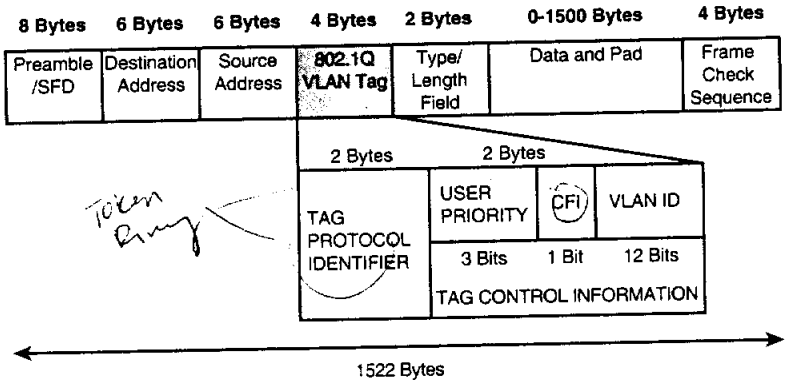


FIGURE 5.11 *The 801.Q VLAN standard uses a frame-tagging approach to identify VLAN membership. The four extra tagging bytes have forced the IEEE to increase the maximum Ethernet frame size to 1522 bytes. This is covered in the 802.3ac standard.*

The extra field has necessitated a maximum frame size increase. The new IEEE 802.3ac standard has increased the maximum Ethernet frame size from 1518 bytes to 1522 bytes.

We can communicate VLAN membership in two ways: explicit and implicit communication. Implicit VLAN membership communication means that the switch or switches know implicitly or indirectly to which particular VLAN a packet belongs. VLAN grouping by IP Subnet addresses is an example of implicit VLAN communication: every packet contains all the information required to identify the VLAN group to which a particular packet belongs.

Explicit membership communication means that the packet or frame needs to be explicitly marked as belonging to a particular VLAN. For example, traffic from a particular MAC address belonging to a specific VLAN would be marked with a specific VLAN identifier.

In general, MAC and port-based VLANs rely on explicit communication, whereas VLANs relying on a Layer 3-based attribute, such as protocol type or Subnet number, can use implicit tagging. Figure 5.12 shows this.

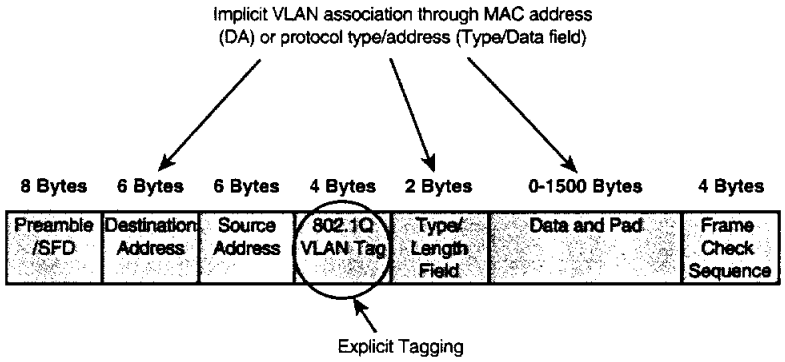


FIGURE 5.12 *VLAN membership can be communicated explicitly or implicitly. First-generation VLANs used only the implicit method: the membership was derived from some other frame variable. The new 802.1Q VLAN tagging standard uses explicit communication.*

VLAN ID numbers need to be centrally assigned and communicated among different switches and nodes on a network; otherwise, the same VLAN ID could exist multiple times. The Generic Attribute Registration Protocol (GARP) is one of the original 802.1P bridging protocol standards. GARP has been used as a foundation for VLAN membership communication among different switches. The GARP VLAN registration protocol (GVRP) assigns and distributes VLAN membership information.

Inter-VLAN Communication

A VLAN is a Layer 2 broadcast domain set up according to a specific grouping parameter. You need to route in order to communicate with another VLAN. There are essentially two ways that routing can occur: either centralized in the form of a classic router or on the edge of the network, for example within a client or server node. The centralized routing is far more popular.

Centralized VLAN Routing

Centralized VLAN routing involves a centralized router that connects the different VLANs at the center of the network, similar to a classic backbone enterprise router.

The so-called "one-armed router" or "router on a stick" attaches to a particular VLAN switch via only a single link. Normally, routers have at least two network links to route among different LANs or a LAN and a WAN link. In this

case, the one-armed router receives the VLAN packet to be routed on its port and performs the address resolution and path calculation. It then sends the packet back with the appropriate routing information so it will reach the correct destination.

Figure 5.13 shows a one-armed router.

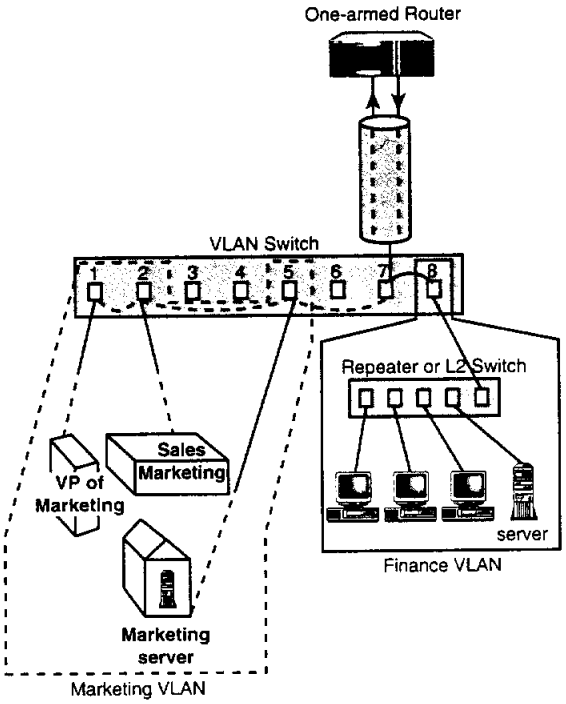


FIGURE 5.13 *A one-armed router can communicate among different VLANs. The term is derived from the fact that this kind of router has only one LAN interface, instead of the customary two.*

Edge Routing

In edge routing, the routing function resides at the perimeter, or edge, of the network and not at the core. For example, a server that is a member of multiple VLANs can perform the routing function. Both NetWare and NT servers can perform the routing function in software. The server could have a dedicated NIC for each VLAN or a single NIC for all VLANs, if the new VLAN tagging standard is supported. If the VLANs are protocol-based, one NIC will also do. Alternate methods will have the clients be part of multiple VLANs from the

start, thereby effectively negating the need for routing. Alternatively, a router could be placed at the edge of the network. Figure 5.14 shows an edge routing example using a server.

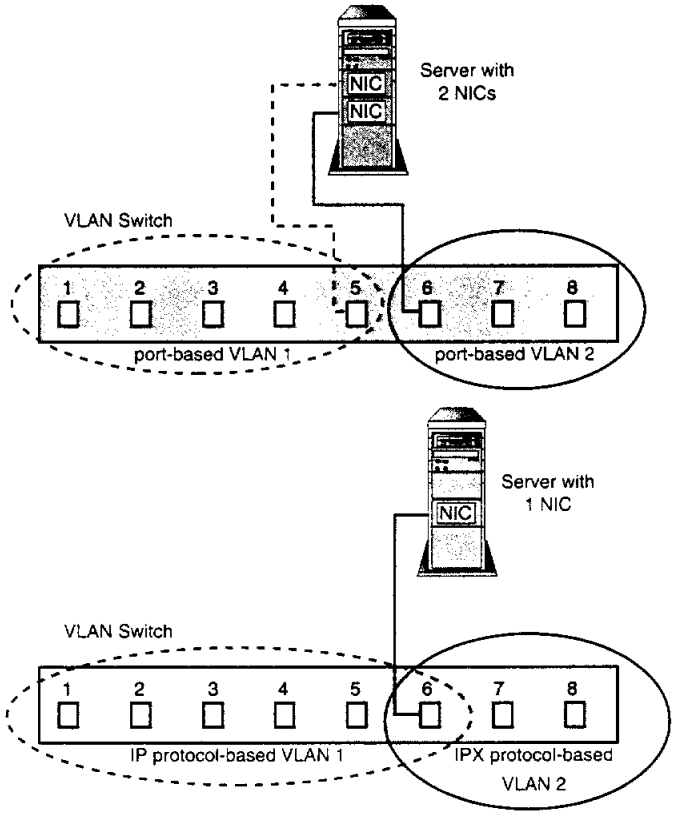


FIGURE 5.14 *Servers can be used as edge routers. Depending on the type of VLAN routing that needs to happen, one or more NICs need to be installed in the server. For example, if the NIC supports the 802.1Q VLAN standard, one NIC is sufficient.*

Layer 3 Switching

The natural evolution of the one-armed router is to integrate both the VLAN switching and routing function into one physical device. Most of the time, this has been accomplished by adding the VLAN function to a high-performance router. The newer method is to add the VLAN and Layer 3 functions to a Layer 2 switch.

Consider Figure 12.12 (a), where a server has been placed on port-based VLAN A. The clients on VLAN B can no longer reach the server because all traffic is filtered between the two VLANs. If all clients need access to this server but you still want to segregate the traffic between VLAN A and B, Figure 12.12 b might provide a better solution. In this figure, all clients can access the server because the switch port it is connected to is *shared* between the two VLANs.

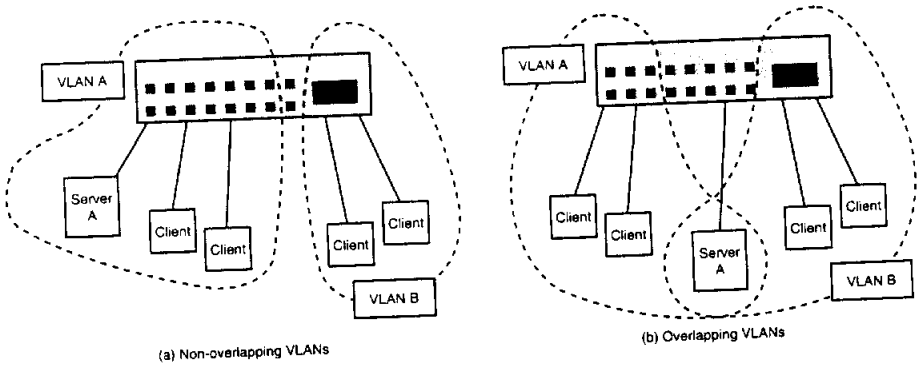


FIGURE 12.12 *A switch with two configured VLANs. VLAN A cannot see traffic from VLAN B, so server A is cut off from clients on VLAN B. The server is connected to a port that is shared between VLAN A and B, so all clients can see the server.*

802.1p Priority Switching

One of ATM's strengths is its capability to prioritize traffic into different classes. Ethernet, on the other hand, has been criticized for its inability to differentiate time-critical data from less important traffic. That's because Ethernet itself is a best-effort data transmission technology, which means there are no service guarantees. Over the last few years, many technologies like 100VG-AnyLAN, 802.3x/BLAM, and 3Com's PACE have tried to improve upon the CSMA/CD access method to improve digital voice and video transmission, but none have been successful to date. With the advent of Fast and Gigabit Ethernet, the overall bandwidth capabilities of Ethernet have vastly improved. Some people argue that full-duplex switched 10, 100, and 1000Mbps Ethernet provides some kind of QoS. This is not true. In effect, full-duplex and switching only increase the bandwidth capabilities of Ethernet. Thus, Ethernet still has no QoS guarantees and no capability to prioritize traffic. (Please refer to Chapter 7 for a general discussion of QoS.)

The new IEEE 802.1Q standard contains a field that allows prioritization of Ethernet traffic. The 3bit field allows you to set eight (that is, 2^3=8) different user priorities for time-critical information.

Note

Sometimes the traffic prioritization standard 802.1p is referred to as 802.1Q/p. Whereas traffic prioritization and the associated protocols are part of the 802.1p spec, the 3bit priority field within an Ethernet frame has been defined within the 802.1Q standard, which also contains the 12-bit VLAN identifier. Thus, priority is really a combination of both 802.1Q and 802.1p, and subsequently, it is often written as the 802.1Q/p standard.

In the past, raw bandwidth was always the cure for delivering time-critical traffic on time. The 802.1Q/p standard now allows manufacturers to build switches and NICs with the inherent capability to prioritize time-critical traffic flows, such as voice or video.

Let's look at an example of how this traffic prioritization works. Figure 5.15 shows two servers that are connected through the same 802.1Q/p-capable switch. The Fileserver is used as a regular file server to deliver large amounts of raw data to a single client (shown as packet stream "FS" in the figure). The Videoserver delivers an IP multicast-based video stream to several nodes. (This is shown as the packet stream labeled "VS.") Switches typically process packets on a FIFO-basis (short for first-in-first-out: The first packet received is also the first one processed for delivery). A problem could arise if the smaller packets of latency-sensitive video traffic (VS) from the server are queued up behind a

large file transfer taking place (FS) at the same time. The video data would be waiting to be sent while the file transfer was taking place. This would result in jitter, which is unacceptable for QoS-sensitive applications, such as video.

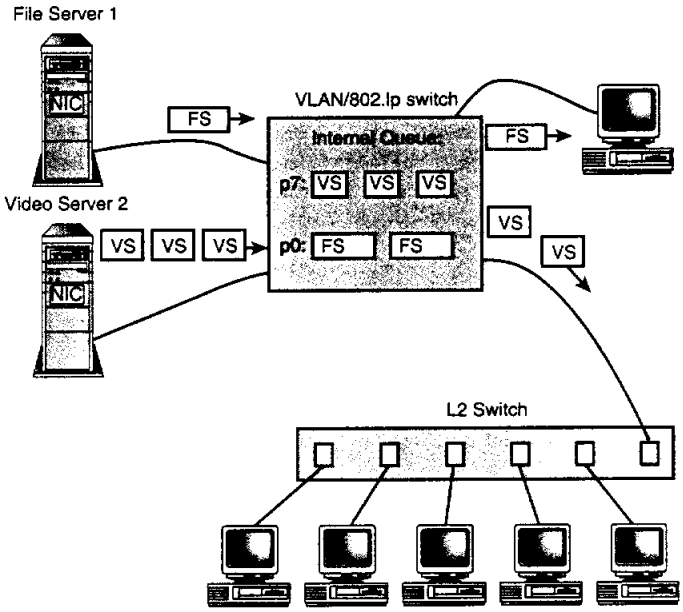


FIGURE 5.15 *The 802.1Q/p priority switching standard ensures on-time delivery of high-priority data and provides Ethernet with QoS capabilities.*

The combination of a server NIC and a switch using the 802.1Q/p priority classification would deliver the video traffic on time, ahead of the file transfer. You would just have to set the file transfer at a lower priority, say priority p1, and set the video stream to a higher priority setting, say priority p7. All switches use a queuing system to analyze and forward traffic. The switch would immediately place incoming frames with priority p7 ahead of the frames with priority p0, ensuring reliable video transmission. For an excellent real-world test of 801.1p-based hardware, take a look at the Tolly Group's test results. Download test report 8283 from `www.tolly.com`.

In our example, both servers communicate through the same VLAN-capable switch. VLANs don't dedicate bandwidth within a switch on their own, so both VLANs are treated equally in the switch.

Note

Providing End-to-End QoS

The IEEE802.1Q/p standard is a Layer 2 traffic classification. The frame priority is typically lost once it reaches a Layer 3 router (unless the router is redesigned or upgraded to understand 802.1p). To provide true end-to-end QoS, you need to use some kind of Layer 3 traffic classification. The Internet Engineering Task Force (IETF) has published an RFC called Resource Reservation Protocol (RSVP) that allows bandwidth to be reserved within a Layer 3 environment to deliver real-time video and audio data streams. The combination of Layer 2/802.1Q/p and RSVP operating at Layer 3 has enabled Ethernet with the capability to compete with ATM in terms of delivering QoS for time-critical applications. The big catch is that existing Ethernet and routing hardware, as well as operating systems and applications, need to be modified to take advantage of these standards. That's going to take a long time.

So Why Haven't VLANs Taken Over the World?

VLANs seem to be very useful entities with numerous benefits. There are, however, many issues that have prevented the widespread adoption of VLANs. Let's look at some of these. In order to be implemented correctly, VLAN nodes require a switched connection. Yes, you can set up VLANs with repeaters "hanging off" switched ports, but in order to implement a VLAN properly, a 100% switched environment is required. This is very rare: Few of today's networks have implemented switching all the way to the desktop.

Technically, VLANs are a Layer 2 function, requiring a router to connect different VLANs. The most effective use of VLANs is probably protocol- or subnet-based VLAN membership. However, a multiport LAN router also can segment a large network into smaller, port-based Subnets, which in some ways accomplishes the same as a VLAN. The high cost and complexity of multiport LAN routers has slowed VLAN adoption. Of course, the new Layer 3 switches discussed in the next section will change all that.

Being a relatively new concept, it will take time for VLANs to be understood, deployed, and managed. You and your networking staff need to learn about these new concepts. Then, of course, come the practical questions, such as what VLAN equipment to buy, how to define membership policies, how to configure VLANs using the new management tools, and so on.

One of the biggest drawbacks to date has been the lack of a distributed VLAN standard. The good news is that the new IEEE 802.1Q VLAN trunking standard will be approved by the time this book is in print, and hardware based on this standard is already shipping.

VLANs group nodes according to specific traffic patterns. Historically, the 80/20 rule has applied to LAN traffic. This means that 80% of traffic is

destined for a node in the local workgroup, whereas 20% is transmitted to a destination outside the workgroup. In a properly defined VLAN environment, this means that 80% of node traffic remains within a VLAN, whereas 20% of traffic is routed to other VLANs or nodes. With the advent of the Web, traffic patterns no longer seem to follow the 80/20 rule. Instead, traffic seems to have taken on more of a 20/80 rule, where 20% of traffic is local and 80% is destined for remote destinations. This is further discussed in Chapter 7 but the implications for VLANs are rather significant. If 80% of traffic leaves a VLAN, why bother at all with grouping nodes according to shared resources? The only benefit of VLANs then becomes broadcast containment. Figure 5.16 illustrates the changing nature of local area traffic flow patterns.

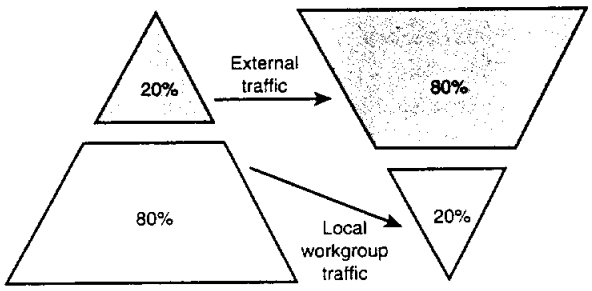


FIGURE 5.16 *Traditionally, 80% of traffic remained in the local workgroup, whereas 20% was external. Recent data shows that this trend is turning upside-down, which has significant implications for VLANs.*

Layer 3 Switching

Now let's discuss Layer 3 switches. We will first look at some of the basics of routing. Then we will look at the evolution of routers, leading up to today's Layer 3 switches. Finally, we will look more closely under the hood of a typical Layer 3 switch.

Simply put, a Layer 3 switch is a different name for a limited purpose, IP/IPX-only router. Depending on what vendor you talk to, Layer 3 switches are also known as routing switches, switch routers, switching routers, wire-speed routers, or hardware-based routers. Taken together, these different names actually describe a Layer 3 switch rather well. We define a Layer 3 switch as a limited-purpose, wire-speed LAN router that uses hardware instead of software to perform packet-to-packet IP routing.

Routing in hardware has two benefits. First, it gives Layer 3 switches the capability to perform routing at wire speeds: such as 10, 100, or even 1000Mbps,

which is something that classic software-based routers couldn't do. Second, hardware-based routers cost significantly less than traditional software-based routers.

A lot of hype has surrounded Layer 3 switches. A word of caution is appropriate at this point. Have you noticed how every hardware manufacturer is hyping Layer 3 switches as though these devices were the answer to all your networking problems? Have you also noticed that Cisco has not been part of this hype frenzy? There's a good reason for this. We all know that Cisco dominates the router market and the high-end backbone LAN router market in particular. These high-end routers, such as the 12,000 and 7X00, are the most profitable pieces of networking hardware equipment sold today, and they provide huge revenues and profits to Cisco. Yet today Cisco's competitors (Bay, 3Com, Cabletron, and new players, such as Extreme, Packet Engines, and Intel) have few alternatives to offer. Therefore, Cisco's competitors have realized that they will never be able to challenge Cisco head-on in the router marketplace. The strategy is simple: If your name isn't Cisco Systems, the only way to get a piece of the backbone action is to change the rules of the game. You have to make believe that routers are obsolete and bet that you can beat Cisco in the new Layer 3 switching game. (Cisco isn't sitting still; it also has a line of Layer 3 switches.)

The analogy is the mainframe and client/server computing. Mainframes were pronounced dead at least a decade ago, yet IBM still derives a very large percentage of its revenues and profits from these heavy iron machines. Mainframes, of course, haven't died yet because they offer a lot of other features that even the best servers to date can't match.

Yes, Layer 3 switches are a technological breakthrough, but don't expect classic multiprotocol chassis routers to disappear overnight.

Layer 3 Background

This isn't a book about routing, so we can't venture too far into the Layer 3 world. The next section will provide some background on Layer 3 protocols. Packet-based Layer 3 routers use standard routing protocols to forward packets. Routing protocols define how routers communicate with each other. Different routers always exchange information that mainly contains route tables, which are similar to the bridge address tables discussed previously, yet they contain a lot more information such as route hops, age, and cost. Most, but not all, networking protocols are routable.

Networking Protocols

Many different Layer 3 networking protocols are in existence today. The most commonly used ones are IP, IPX, SNA, NetBIOS, NetBEUI, AppleTalk, and DECnet. To us, this sounds like *NIH* (not invented here). Historically, individual networking software or hardware vendors needed to invent their own protocols. Consequently, dozens of different protocols are in use to this date. For example, Novell decided the Internet Protocol wasn't good enough and invented Extended IP. Most protocols in use today are LAN focused. They aren't routable and make extensive use of broadcasts to advertise the existence of servers on the LAN. All protocols have their advantages and disadvantages, which in a way is a moot point, as IP is fast becoming the de facto standard.

Routable Protocols

Only IP, IPX, DECnet, and AppleTalk protocols are routable. Other nonroutable protocols can be routed through a technology called IP Tunneling, where the nonroutable packets are transported in IP packets. If a protocol cannot be routed, it must be bridged. Table 5.2 summarizes the most popular networking protocols today.

TABLE 5.2 SUMMARY OF COMMON NETWORKING PROTOCOLS

Protocol	Short for	Vendor	Products using	Routable?
IP	Internet Protocol	IETF Open Standard	All	Yes
IPX	Internetwork Protocol Exchange	Novell	NetWare	Yes
AppleTalk	Apples talking to each other	Apple	Macintosh PCs and Servers	Yes
NetBIOS	Network Basic Input/Output System	Microsoft, IBM	Microsoft Windows, Microsoft LAN Manager, IBM LANServer	No
NetBEUI	NetBIOS Extended User Interface	Microsoft, IBM	Same as NetBIOS	No
SNA	Systems Network Architecture	IBM	Legacy Mainframes	No
DECnet	DEC[1] Network	Compaq[1]	Legacy Alpha/VAX	Yes
LAT	Local area transport	Compaq	Legacy Alpha/VAX	No

[1] *Compaq Computer acquired Digital Equipment Corporation in 1998.*

Note

The Internet Engineering Task Force (IETF) is the definitive standards authority for Layer 3 and 4 technologies. Essentially, the IETF is to Layer 3 and 4 what the IEEE is to Layer 2 and Ethernet. The IETF publishes Request For Comments (or RFC) standards documents. The term RFC is a little misleading. There are draft RFCs, and official standards RFCs. Drafts are documents circulated for discussion, and approved IETF RFCs become Internet routing standards.

Routing Protocols

Different protocols require different routing protocols, too. The most common IP routing protocols are RIP, RIP2, and OSPF. IPX uses the RIP protocols too, but it's slightly different from the IP RIP protocol.

RIP (also known as RIP1 or RIPv1) stands for *routing information protocol* and was the first industry standard IP routing standard. It was initially developed for the BSD version of UNIX but became an industry standard over time. RIP2 is an improved version of the original RIP1 protocol in one sense because it uses fewer broadcasts.

OSPF stands for *open shortest path first* and is a newer IP development. OSPF has many advantages over RIP. Layer 2 allows for only one active path. The spanning tree algorithm shuts down parallel paths. Routed networks, on the other hand, encourage multiple redundant paths. OSPF allows routers to determine the lowest cost and shortest path from different choices more accurately. If the active path fails, OSPF will change the route much more quickly than RIP. OSPF also doesn't send out route tables like RIP does, reducing broadcast traffic rates. For LAN routing, the added benefit of OSPF does not always outweigh its added complexity.

Over the years, several multicast routing protocols have also been developed. Table 5.3 shows the most common routing protocols in use.

TABLE 5.3 SUMMARY OF COMMON IP AND IPX ROUTING PROTOCOLS

Protocol Supported	Routing Protocol	Short For	Standard
IP	RIP, RIP2	Routing Information Protocol, Routing Information Protocol 2	RFC 950/1058; RFC 173
IP	OSPF	Open Shortest Path First	RFC 1583, RFC1850
IP	DVMRP, IGMP	Distance Vector Multicast Routing Protocol, Internet Group Membership Protocol	IETF RFC 1162

Protocol Supported	Routing Protocol	Short For	Standard
IP	IGRP, IGRP	Interior Gateway Routing Protocol, Enhanced IGRP	Cisco proprietary
IPX	RIP, SAP, NLSP	Routing Information Protocol, Service Advertising Protocol, NetWare Link State Advertising Protocol	Novell proprietary

How Routers Work

Routers are used in various places in today's networks and form the core of many large networks today.

A detailed discussion of routers is really beyond the scope of this chapter and this book. Here is a quick overview of how routers work.

Figure 5.17 shows a router flow diagram. (Feel free to page back and contrast this diagram with the flow diagram for a Layer 2 bridge shown in Figure 4.4.)

A router performs the following functions:

- A received frame is stored and stripped of all its Layer 2 header information. The exposed *data* field then contains the actual Layer 3 packet. This means that the router needs to discard the Ethernet SA, DA, and so on.
- The protocol type is identified. With some Ethernet frame types, the header will contain this information, but in general, the actual protocol can only be identified by looking at the data field itself.
- The router determines whether the protocol is routable. If it is not, the packet is reassembled into its original Layer 2 frame and switched or bridged. (Refer to Chapter 4 for details.)
- If the protocol is routable (IP, IPX, AppleTalk, or DECnet), the router will perform some housekeeping items like check the hopcount number. A maximum of 15 router hops are typically allowed, after which a packet is discarded. This prevents packets from endlessly being routed looking for a destination that might not exist. The hop counter is then incremented.
- The router will look up the destination IP address in its routing table and identify the relevant destination port. If the destination node is directly attached to the router (such as another port on the router), the relevant

Layer 2 information, such as the original MAC address, is added back and the frame forwarded.

- If the destination node is not directly attached (another router hop needs to occur), the relevant IP address is added to the packet. Then the Layer 2 information, including new MAC address from the routing table, are added back again before the frame is forwarded to the relevant port.

In general, a router functions similarly to a bridge in that it looks up the destination address in a table. The key differences between bridges and routers occur in how the address table, or the route, is communicated among different routers. Whereas switches and bridges act as standalone devices (they create their own address tables, based on SA and DA information learned from frames received), a router obtains its table information through both learning and exchanging routing tables with other routers. A router is, in essence, part of a larger routing fabric that knows the exact traffic route spanning multiple routers.

Historically, routers were used as remote access devices to link different LANs together over remote distances. The Internet, for example, is essentially a worldwide maze of different interconnected IP routers.

Routers can translate from one protocol to another. For example, if one LAN uses the Novell IPX protocol and a second network utilizes IP, a router is required to translate between the two. As the number of different networking protocols increased, routers began to take on a different role within the network, as opposed to the edge of the network.

When customers implemented new technologies, such as FDDI, routing needed to take place in order to convert traffic from Ethernet to the new FDDI frame format.

As LANs began to grow in size and traffic volume, routers migrated to the center of the network to form a central backbone. This provides several benefits:

- The WAN access point is now centrally located.
- The protocol conversion happens at the most efficient point, namely in the center.
- Dividing up a large flat Layer 2 network into multiple Layer 3 broadcast domains reduces the network utilization, boosting performance.
- Multiple subnets made the network less susceptible to broadcast storms.
- Different VLANs needed to communicate via a router.

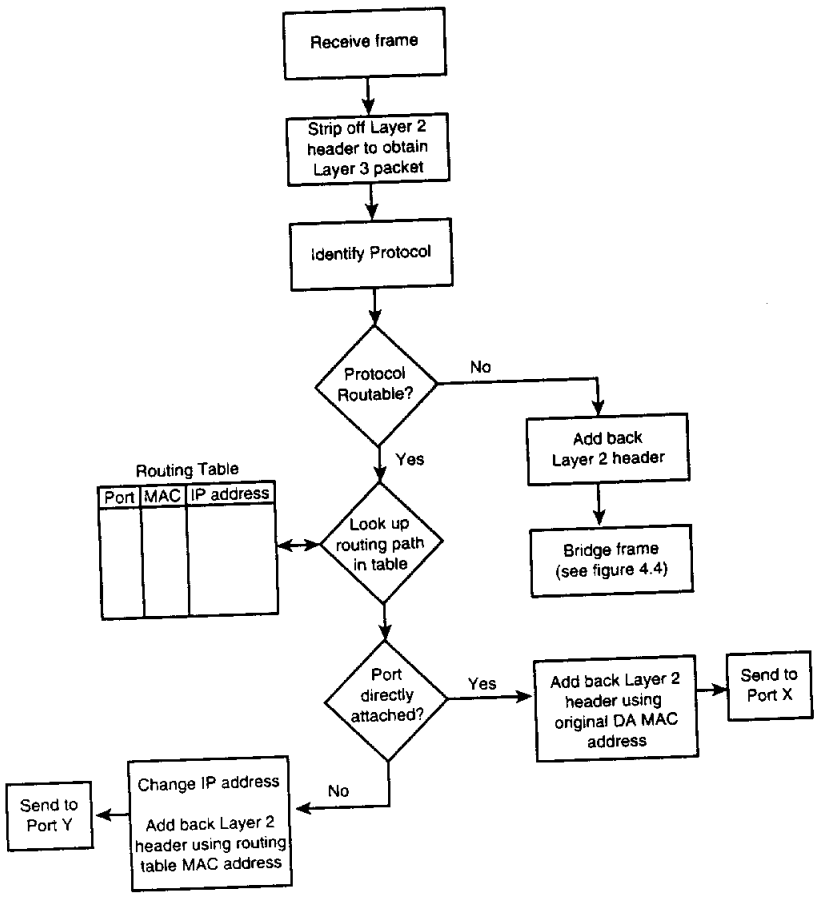


FIGURE 5.17 *Router internal forwarding logic looks similar to a bridge flow diagram shown in Figure 4.4. It does contain an added branch that performs the routing function.*

Ultimately, large backbone routers made a network more stable, reliable, and manageable.

What Exactly Is a Layer 3 Switch?

Layer 3 switches are limited-purpose, hardware-based IP routers. The term *Layer 3 switch* is a marketing creation, just like Ethernet Layer 2 switches are all really just multiport bridges. So, technically, IEEE Layer 2 bridging and IETF Layer 3 routing standards govern Layer 3 switches.

Layer 3 switches do the following:

- They operate primarily at Layer 3 of the OSI model but also perform frame forwarding at Layer 2.
- They only route IP or IPX protocols (see Table 5.4). Packet routing is done in accordance with established Layer 3 routing standards, for example through the exchange of routing tables based on industry standard routing protocols, such as RIP. (See the section "Router Protocols" earlier in this chapter.)
- They switch nonroutable traffic at Layer 2.
- They forward frames at wire speed rates, such as 10, 100, or 1000Mbps, with minimal latencies, typically a few microseconds. (See the section "So How Fast Is Wire Speed?" later in this chapter.)
- They support only LAN-based routing. Right now, only Ethernet Layer 3 switches are available. Someone might very likely build a Token Ring Layer or FDDI Layer 3 switch one day.
- They are substantially cheaper than similar performance routers.

Table 5.4 compares classical LAN routers with Layer 3 switches.

TABLE 5.4 A COMPARISON OF LAYER 3 SWITCHES AND ROUTERS

Feature	Classical LAN Router	Layer 3 Switch
OSI layer function performed	Layer 3	Layer 3
Routing performed	Software (CPU plus software)	Hardware (ASIC chips)
Layer 2 MAC supported	Ethernet, Token Ring, FDDI, ATM, WAN	Fast and Gigabit Ethernet only (so far)
Performance	Slow to medium, depending on CPU performance and cost	Fast, near wire speed
Price/port	High	Low
Latency	Approximately 200µs	<10µs (100Mbps)
Programmability and manageability	Extremely high	Little to none
Protocols supported	All	IP, sometimes IPX[1]
Routing protocols proprietary	All	RIP1, RIP2, sometimes OSPF and DVMRP[2]

Feature	Classical LAN Router	Layer 3 Switch
Application	Creating broadcast domains via a collapsed or distributed backbone	Most places where a Layer 2 switch is used today
	WAN connection	Collapsed backbone
	Multiprotocol routing	Inter-VLAN routing

[1] *Some vendors are talking about embedding additional protocols in hardware. The two that are being mentioned are AppleTalk and DECnet, both of which are routable but have relatively small and shrinking market share. Over time, more protocols will be added into hardware, and, after IPX, these two are next.*

[2] *Just like the number of protocols embedded in hardware is bound to increase, so is the number of routing protocols. Right now, most vendors support RIP and RIP 2. One or two switches feature OSPF and DVMPP. The list of protocols is bound to grow over time.*

So How Fast Is Wire Speed?

Wire speed can be measured in Megabits-per-second or packets-per-second. Layer 3 devices are measured in packets-per-second, not frames-per-second. That's because routers operate at Layer 3, where packets are the norm, as opposed to Layer 2, where frames are the norm. A Layer 3 packet is always encapsulated within a Layer 2 frame, so packet and frame rates are the same. The frame is a little larger than a packet, because it contains some additional bytes like SA, DA, and preamble. In this context, you can use frame and packet interchangeably. Sometimes, pps is even used for Layer 2 devices, although that is technically not correct. This might occur to avoid confusion with video transmission, where frames-per-second is the prevalent way of measuring video transfer rates.

When comparing the performance of Ethernet Layer 3 switches, you need to know what packet size was used to measure the speed. Most labs stress-test switches for different packet sizes, ranging from the minimum to the maximum packet sizes. The minimum packet size of 64 bytes will result in the maximum packet rate, which is 100Mbps Ethernet 148,810 packets per second (fps) to be exact. This will stress a switch to the limit, as the switch will have to perform an IP routing function for every packet forwarded. For the maximum packet size of 1518 bytes, the packet rate is 8127 fps for 100Mbps Ethernet. For 10Mbps Ethernet, these numbers are one-tenth; for Gigabit Ethernet, the maximum packet rate can be an astounding 1.48 million packets per second.

Table 5.5 shows the calculation of the maximum packet rates. Note that many tests show packets-per-second, which is confusing, as Ethernet operates at Layer 2, where the term *frames* is used. This can become very confusing, especially when testing Layer 3 devices, where both frame and packet rates are applicable.

TABLE 5.5 100MBPS ETHERNET WIRE SPEED FRAME RATES

Packet Size[1] (bytes)	Preamble + IFG[2] (bytes)	Packet + Preamble + IFG[3] (bytes)	Total Packet Size[4] (bits)	Packet Length[5] (µs)	Maximum Rate[6] (pps)
64	20	84	672	6.72	148,810
1518	20	1538	12,304	120	8127

1 *Minimum and maximum packet size in bytes is shown.*
2 *Every Ethernet frame has a Preamble, which is 8 bytes. IFG is the minimum Ethernet Interframe Gap that separates subsequent frame transmissions and is 12 bytes long.*
3 *Packet size added to (Preamble + IFG).*
4 *Packet size in bits is (frame + preamble + IFG)×8 bits.*
5 *Bit-time in nanoseconds = 1/wire speed. For 100Mbps Ethernet, bit-time = 1/[100×10E+06] = 10ns. packet length = bit-time×packet size.*
6 *Maximum packet rate in packets-per-second (pps) = wire speed/packet length.*

The latency of Layer 3 switches is typically a few microseconds. The relevant IP information needs to be decoded before the frame can be forwarded. Most Layer 3 switches allow you to select store-and-forward or cut-through. If the switch operates in store-and-forward mode, the entire frame needs to be received, in which case the latencies can be as large as the longest frames encountered. In cut-through mode, the latency can be as low as a few hundred bit-times. For 100Mbps switches, this translates to a few microseconds.

Market Trends

Let's look at some recent market trends that have enabled the emergence of Layer 3 switches:

- *Switching is the LAN buzzword of the 90s*—Over the past five years, Layer 2 Ethernet switch sales have exploded, driven by intense competition and declining prices. Workgroup switches connect users at 10Mbps, servers and uplinks connect to 100Mbps switches, and then Gigabit Ethernet backbone switches tie it all together. The vision is simple: Switching everywhere.

  Of course, the ATM hype helped fuel this one, too: Shared Ethernet, and to a lesser degree, routing, are no longer politically correct. We hear someone disagreeing in the background: "Houston, we have a problem." We can't really live without routers yet. So some clever marketing people read about the old Kalpana idea of calling a bridge a switch. Why not just call the new router a switch, too? In order to differentiate these switching routers from classic bridging switches, something else was needed: Add the word *Layer 3* and *BINGO*. We have a new product category: the Layer 3 switch.

- *ASIC capabilities have increased exponentially*—Gordon Moore, cofounder of Intel, observed about 20 years ago that the transistor density on an integrated circuit can double every 24 months. While this so-called Moore's law is well-known in the PC industry (see Chapter 7), this exponential increase in circuit complexity allows very sophisticated chips to be designed for other functions, too.

  Layer 3 switches often use a type of chip known as an *ASIC*, short for *application-specific integrated circuit*. ASICs are nothing new: They have been around for at least a decade (although some switch vendors pretend like they personally invented the ASIC only yesterday, specifically for their particular switch). ASICs are high-density integrated circuits designed for a specific task only, to be produced in a relatively short time and in smaller numbers, with limited complexity. Custom chips, on the other hand, take longer to design, are more complex, and require higher run rates in order to recover the more expensive and time-consuming up-front development. General-purpose integrated circuits, such as microprocessors, memory chips, Ethernet MAC, and PHYs, are all custom chips. ASICs have grown in complexity over the last few years, due to Moore's law and advances in ASIC design tools. As a result, both Layer 2 and Layer 3 routing function can now be hardwired into an integrated circuit chip, as opposed to being performed by a software program.

- *The undisputed winner of the protocol wars is IP*—A few years ago, numerous networking protocols existed side-by-side with IPX dominating in the PC world and SNA in the mainframe world. There were, of course, other protocols from Microsoft, DEC, Banyan, and IBM. Then there was the UNIX world and the Internet, where TCP/IP ruled. With the explosion of the Internet, IP has emerged as the de facto standard. Building a multiprotocol router in hardware would technically be possible, but the complexity of such a design would probably mean a development effort costing far too much money for anyone to attempt. This standardization on IP has made it possible to build IP and only IP routing into Layer 3 switches. (Every new protocol to be routed would increase the complexity of the hardware tremendously.)

- *IP is mature and stable*—One of the advantages of classic software-based routing is that routers are reprogrammable. If a bug emerges in the routing software or the protocol standard evolves, you can merely download a new version of the relevant routing software and upgrade the router software, which is typically stored in Flash memory. Putting the routing logic in hardware has tremendous speed advantages but locks you into

the same logic for life! Imagine buying a few new chips and hauling out your soldering iron every time you want to upgrade your router software! The good news is that IP and associated IP routing standards, such as RIP or RIP2, have matured tremendously over the last few years. Today, IP is so stable and has proven that it lends itself to being embedded in hardware.

The bottom line is that semiconductor advances and the de facto standardization of IP have enabled the emergence of Layer 3 switches.