This chapter covers the following topics:

- L2 Switching Basics
- Metro Ethernet Services Concepts
- Example of an L2 Metro Ethernet Service
- Challenges with All-Ethernet Metro Networks

CHAPTER 3

Metro Ethernet Services

As discussed in Chapter 1, "Introduction to Data in the Metro," Ethernet services can take either of two forms: a retail service that competes with traditional T1/E1 private-line services, or a wholesale service where a carrier sells a big Ethernet transport pipe to another, smaller service provider. In either case, multiple customers share the same metro infrastructure and equipment. For TDM deployments, sharing the infrastructure is a nonissue, because the services are limited to selling transport pipes, and each customer is allocated a circuit that isolates its traffic from other customers. The customer gets well-defined SLAs, mainly dictated by the circuit that is purchased.

When packet multiplexing and switching are applied, such as in the cases of switched EOS, Ethernet Transport, and RPR, things change. Packets from different customers are multiplexed over the same pipe, and the bandwidth is shared. No physical boundaries separate one customer's traffic from another's, only logical boundaries. Separation of customer traffic and packet queuing techniques have to be used to ensure QoS. Multiple functions have to be well-defined to offer a service:

- How to identify different customers' traffic over a shared pipe or shared network
- How to identify and enforce the service given to a particular customer
- How to allocate certain bandwidth to a specific customer
- How to "transparently" move customers' traffic between different locations, such as in the case of transparent LAN services
- How to scale the number of customers
- How to deploy a VPN service that offers any-to-any connectivity for the same customer

This chapter starts by discussing the basics of L2 Ethernet switching to familiarize you with Ethernet-switching concepts. Then it discusses the different metro Ethernet service concepts as introduced by the Metro Ethernet Forum (MEF).

L2 Switching Basics

L2 switching allows packets to be switched in the network based on their Media Access Control (MAC) address. When a packet arrives at the switch, the switch checks the packet's destination MAC address and, if known, sends the packet to the output port from which it learned the destination MAC.

The two fundamental elements in Ethernet L2 switching are the MAC address and the virtual LAN (VLAN). In the same way that IP routing references stations on the networks via an L3 IP address, Ethernet L2 switching references end stations via the MAC address. However, unlike IP, in which IP addresses are assigned by administrators and can be reused in different private networks, MAC addresses are supposed to be unique, because they are indicative of the hardware itself. Thus, MAC addresses should not be assigned by the network administrator. (Of course, in some cases the MAC addresses can be overwritten or duplicated, but this is not the norm.)

Ethernet is a broadcast medium. Without the concept of VLANs, a broadcast sent by a station on the LAN is sent to all physical segments of the switched LAN. The VLAN concept allows the segmentation of the LAN into logical entities, and traffic is localized within those logical entities. For example, a university campus can be allocated multiple VLANs—one dedicated for faculty, one dedicated for students, and the third dedicated for visitors. Broadcast traffic within each of these VLANs is isolated to that VLAN.

Figure 3-1 shows the concept of an Ethernet LAN using a hub (Part A) and an Ethernet switch (Part B). With an Ethernet hub, all stations on the LAN share the same physical segment. A 10-Mbps hub, for example, allows broadcast and unicast traffic between the stations that share the 10-Mbps bandwidth. The switched LAN on the right allows each segment a 100-Mbps connection (for this example), and it segments the LAN into two logical domains, VLAN 10 and VLAN 20. The concept of VLANs is independent of the stations themselves. The VLAN is an allocation by the switch. In this example, ports 1 and 2 are allocated to VLAN 10, while ports 3 and 4 are allocated to VLAN 20. When stations A1 and A2 send traffic, the switch tags the traffic with the VLAN assigned to the interface and makes the switching decisions based on that VLAN number. The result is that traffic within a VLAN is isolated from traffic within other VLANs.

Ethernet switching includes the following basic concepts:

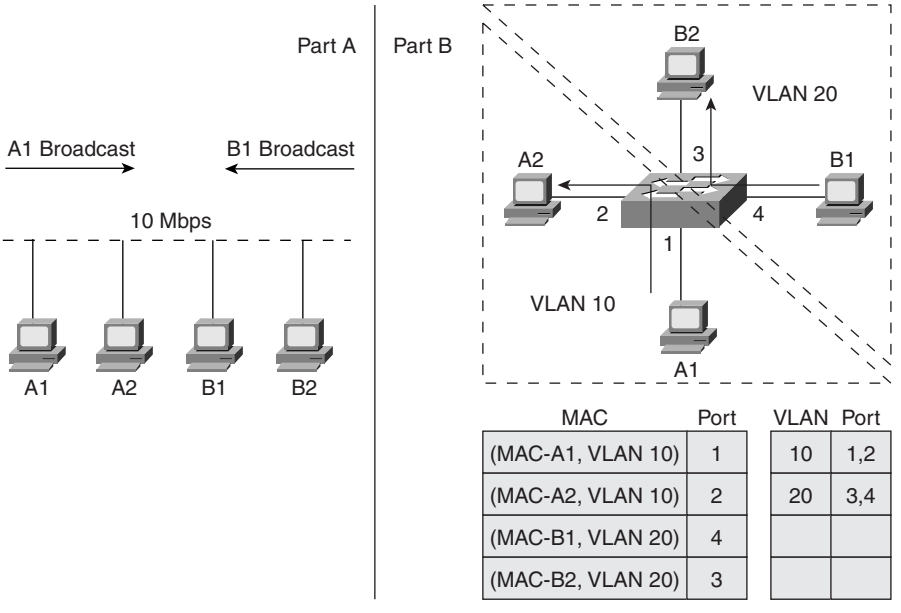
- MAC learning
- Flooding
- Using broadcast and multicast
- Expanding the network with trunks
- VLAN tagging
- The need for the Spanning Tree Protocol (STP)

MAC Learning

MAC learning allows the Ethernet switch to learn the MAC addresses of the stations in the network to identify on which port to send the traffic. LAN switches normally keep a MAC learning table (or a bridge table) and a VLAN table. The MAC learning table associates the MACs/VLANs with a given port, and the VLAN table associates the port with a VLAN. In Figure 3-1, Part B, the switch has learned the MAC addresses of stations A1, A2, B1, and B2

on ports 1, 2, 4, and 3, respectively. It also shows that ports 1 and 2 are associated with VLAN 10 and ports 3 and 4 are associated with VLAN 20.

Figure 3-1 *Ethernet MACs and VLANs*



Flooding

If the switch receives a packet with a destination MAC address that does not exist in the bridge table, the switch sends that packet over all its interfaces that belong to the same VLAN assigned to the interface where the packet came in from. The switch does not flood the frame out the port that generated the original frame. This mechanism is called *flooding*. It allows the fast delivery of packets to their destinations even before all MAC addresses have been learned by all switches in the network. The drawback of flooding is that it consumes switch and network resources that otherwise wouldn't have been used if the switch had already learned which port to send the packet to.

VLANs minimize the effect of flooding because they concentrate the flooding within a particular VLAN. The switch uses the VLAN table to map the VLAN number of the port on which the packet arrived to a list of ports that the packet is flooded on.

Using Broadcast and Multicast

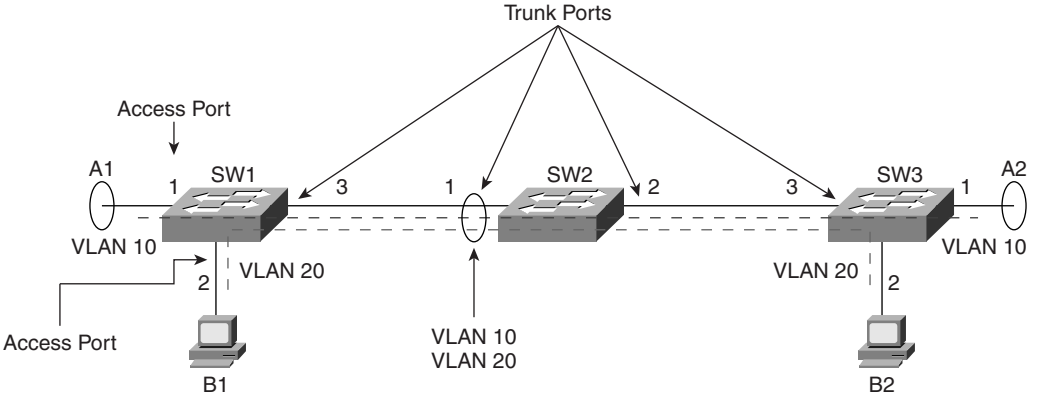
Broadcast is used to enable clients to discover resources that are advertised by servers. When a server advertises its services to its clients, it sends broadcast messages to MAC address FFFF FFFF FFFF, which means "all stations." End clients listen to the broadcast and pick up

only the broadcasts they are interested in, to minimize their CPU usage. With multicast, a subset of broadcast, a station sends traffic only to a group of stations and not to all stations. Broadcast and multicast addresses are treated as unknown destinations and are flooded over all ports within a VLAN. Some higher-layer protocols such IGMP snooping help mitigate the flooding of IP multicast packets over an L2 switched network by identifying which set of ports a packet is to be flooded on.

Expanding the Network with Trunks

So far you have seen the case of a single L2 switch. An L2 Ethernet-switched network would consist of many interconnected switches with trunk ports. The trunk ports are similar to the access ports used to connect end stations; however, they have the added task of carrying traffic coming in from many VLANs in the network. This scenario is shown in Figure 3-2. Trunk ports could connect Ethernet switches built by different vendors—hence the need for standardization for VLAN tagging mechanisms.

Figure 3-2 *Trunk Ports*

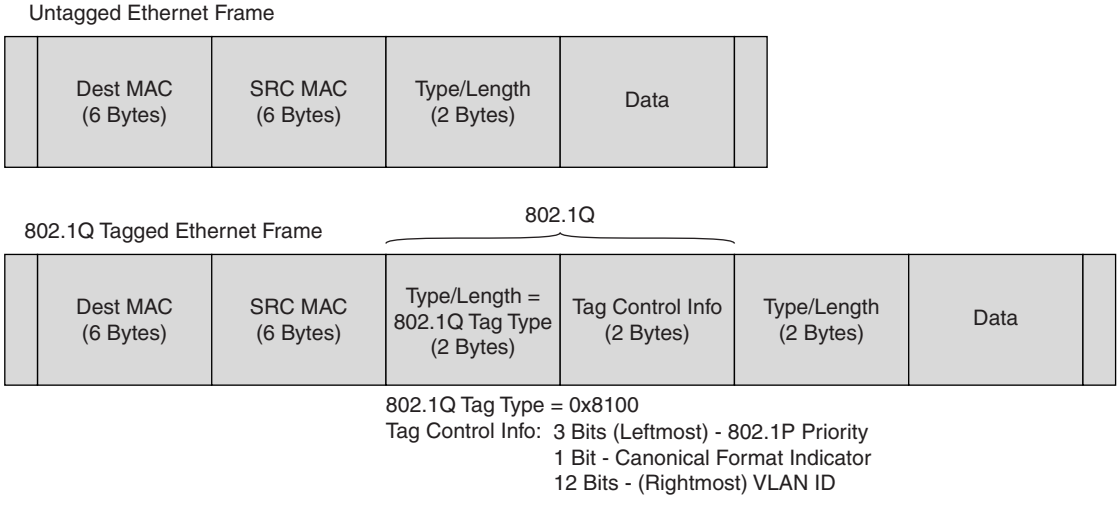


In Figure 3-2, switches SW1 and SW3 have assigned access port 1 with VLAN 10 and access port 2 with VLAN 20. Port 3 is a trunk port that is used to connect to other switches in the network. Note that SW2 in the middle has no access ports and is used only to interconnect trunk ports. You can see that the simplicity of switched Ethernet becomes extremely complex because VLAN assignments need to be tracked inside the network to allow the right traffic to be switched on the right ports. In Frame Relay, ATM, and MPLS, similar complexities do exist, and signaling is introduced to solve the network connectivity issues. Ethernet has *not* defined a signaling protocol. The only mechanisms that Ethernet networks have are third-party applications that surf the network and make it easier to do some VLAN allocations. While these mechanisms work in small enterprise environments, they immediately became showstoppers in larger enterprise deployments and carrier networks. Chapter 4 discusses LDP as a signaling mechanism for delivering Ethernet services. Chapter 7 discusses RSVP-TE and its use in relation to scaling the Ethernet services.

VLAN Tagging

IEEE 802.1Q defines how an Ethernet frame gets tagged with a VLAN ID. The VLAN ID is assigned by the switch and not the end station. The switch assigns a VLAN number to a port, and every packet received on that port gets allocated that VLAN ID. The Ethernet switches switch packets between the same VLANs. Traffic between different VLANs is sent to a routing function within the switch itself (if the switch supports L3 forwarding) or an external router. Figure 3-3 shows how the VLAN tags get inserted inside the untagged VLAN packet.

Figure 3-3 *VLAN Tagged Packet*



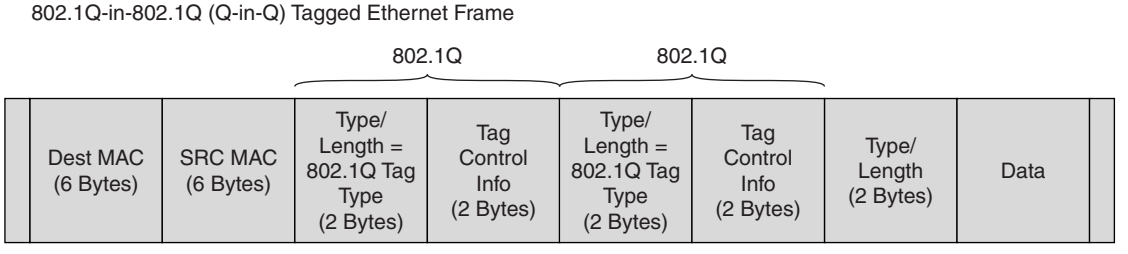
The untagged Ethernet packet consists of the destination MAC address and source MAC address, a Type field, and the data. The 802.1Q tag header gets inserted between the source MAC address and the Type field. It consists of a 2-byte Type field and a 2-byte Tag Control Info field. The 2-byte Type field is set to 0X8100 to indicate an 802.1Q tagged packet. The 2-byte Tag Control Info field consists of the 3 leftmost bits indicating the 802.1P priority and the 12 rightmost bits indicating the VLAN tag ID. The 802.1P field gives the Ethernet packet up to eight different priority levels that can be used to offer different levels of service within the network. The 12-bit VLAN ID field allows the assignment of up to 4096 (2^{12}) VLAN numbers to distinguish the different VLAN tagged packets.

Metro Ethernet applications require extensions to L2 switching that are not defined in the standards. An example is the ability to do VLAN stacking—that is, to do multiple VLAN tagging to the same Ethernet packet and create a stack of VLAN IDs. Different entities can do L2 switching on the different levels of the VLAN stack. Cisco Systems calls this concept *Q-in-Q,* short for *802.1Q-in-802.1Q,* as shown in Figure 3-4.

As shown, an already tagged frame can be tagged again to create a hierarchy. The simplicity of Ethernet, the lack of standardization for many such extensions, the reliance on STP, and

the explosion of MAC addresses contribute to the lack of confidence of many providers in deploying a large-scale, all-Ethernet network.

Figure 3-4 *Q-in-Q*



VLAN tag support is discussed more in the section "VLAN Tag Support Attribute."

The Need for the Spanning Tree Protocol

L2 Ethernet-switched networks work on the basis of MAC address learning and flooding. If multiple paths exist to the same destination, and the packet has an unknown destination, packet flooding might cause the packet to be sent back to the original switch that put it on the network, causing a broadcast storm. STP prevents loops in the network by blocking redundant paths and ensuring that only one active path exists between every two switches in the network. STP uses bridge protocol data units (BPDUs), which are control packets that travel in the network and identify which path, and hence ports, need to be blocked.

The next section covers in detail the Ethernet services concepts as defined by the Metro Ethernet Forum.

Metro Ethernet Services Concepts

The Metro Ethernet Forum is a nonprofit organization that has been active in defining the scope, concepts, and terminology for deploying Ethernet services in the metro. Other standards bodies, such as the Internet Engineering Task Force (IETF), have also defined ways of scaling Ethernet services through the use of MPLS. While the terminologies might differ slightly, the concepts and directions taken by these different bodies are converging.

For Ethernet services, the MEF defines a set of attributes and parameters that describe the service and SLA that are set between the metro carrier and its customer.

Ethernet Service Definition

The MEF defines a User-to-Network Interface (UNI) and Ethernet Virtual Connection (EVC). The UNI is a standard Ethernet interface that is the point of demarcation between the customer equipment and the service provider's metro Ethernet network.

The EVC is defined by the MEF as "an association of two or more UNIs." In other words, the EVC is a logical tunnel that connects two (P2P) or more (MP2MP) sites, enabling the transfer of Ethernet frames between them. The EVC also acts as a separation between the different customers and provides data privacy and security similar to Frame Relay or ATM permanent virtual circuits (PVCs).

The MEF has defined two Ethernet service types:

- **Ethernet Line Service (ELS)**—This is basically a point-to-point (P2P) Ethernet service.
- **Ethernet LAN Service (E-LAN)**—This is a multipoint-to-multipoint (MP2MP) Ethernet service.

The Ethernet Line Service provides a P2P EVC between two subscribers, similar to a Frame Relay or private leased-line service (see Figure 3-5).

Figure 3-5 *Ethernet Service Concepts*

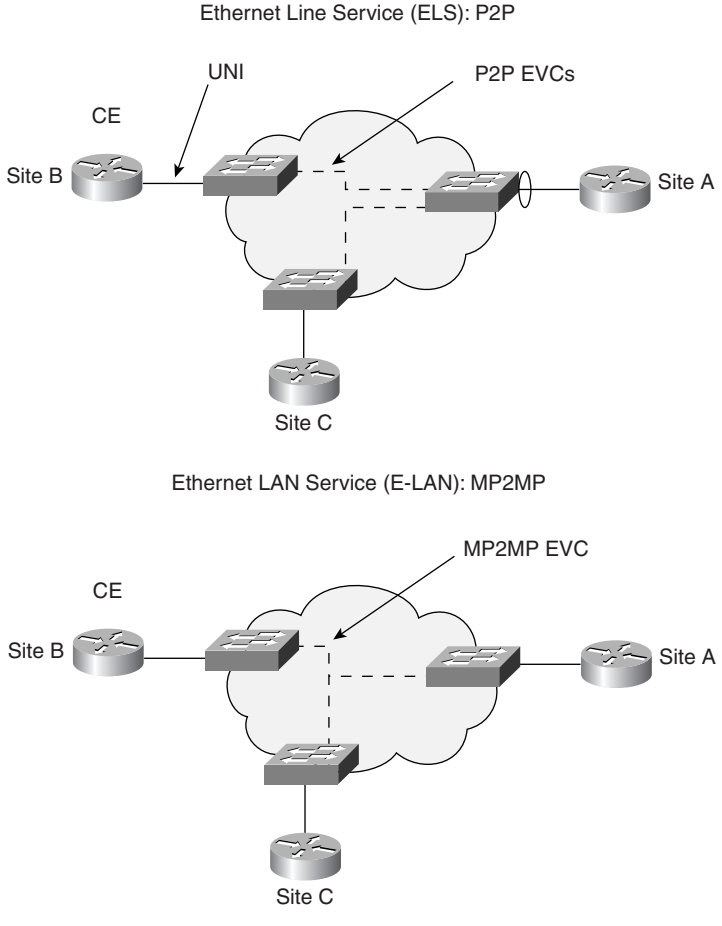


Figure 3-5 also illustrates the E-LAN, which provides multipoint connectivity between multiple subscribers in exactly the same manner as an Ethernet-switched network. An E-LAN service offers the most flexibility in providing a VPN service because one EVC touches all sites. If a new site is added to the VPN, the new site participates in the EVC and has automatic connectivity to all other sites.

Ethernet Service Attributes and Parameters

The MEF has developed an Ethernet services framework to help subscribers and service providers have a common nomenclature when talking about the different service types and their attributes. For each of the two service types, ELS and E-LAN, the MEF has defined the following service attributes and their corresponding parameters that define the capabilities of the service type:

- Ethernet physical interface attribute
- Traffic parameters
- Performance parameters
- Class of service parameters
- Service frame delivery attribute
- VLAN tag support attribute
- Service multiplexing attribute
- Bundling attribute
- Security filters attribute

Ethernet Physical Interface Attribute

The Ethernet physical interface attribute has the following parameters:

- **Physical medium**—Defines the physical medium per the IEEE 802.3 standard. Examples are 10BASE-T, 100BASE-T, and 1000BASE-X.
- **Speed**—Defines the Ethernet speed: 10 Mbps, 100 Mbps, 1 Gbps, or 10 Gbps.
- **Mode**—Indicates support for full duplex or half duplex and support for autospeed negotiation between Ethernet ports.
- **MAC layer**—Specifies which MAC layer is supported as specified in the 802.3-2002 standard.

Traffic Parameters

The MEF has defined a set of bandwidth profiles that can be applied at the UNI or to an EVC. A bandwidth profile is a limit on the rate at which Ethernet frames can traverse the UNI or the

EVC. Administering the bandwidth profiles can be a tricky business. For P2P connections where there is a single EVC between two sites, it is easy to calculate a bandwidth profile coming in and out of the pipe. However, for the cases where a multipoint service is delivered and there is the possibility of having multiple EVCs on the same physical interface, it becomes difficult to determine the bandwidth profile of an EVC. In such cases, limiting the bandwidth profile per UNI might be more practical.

The Bandwidth Profile service attributes are as follows:

- Ingress and egress bandwidth profile per UNI
- Ingress and egress bandwidth profile per EVC
- Ingress and egress bandwidth profile per CoS identifier
- Ingress bandwidth profile per destination UNI per EVC
- Egress bandwidth profile per source UNI per EVC

The Bandwidth Profile service attributes consist of the following traffic parameters:

- **CIR (Committed Information Rate)**—This is the minimum guaranteed throughput that the network must deliver for the service under normal operating conditions. A service can support a CIR per VLAN on the UNI interface; however, the sum of all CIRs should not exceed the physical port speed. The CIR has an additional parameter associated with it called the Committed Burst Size (CBS). The CBS is the size up to which subscriber traffic is allowed to burst in profile and not be discarded or shaped. The in-profile frames are those that meet the CIR and CBS parameters. The CBS may be specified in KB or MB. If, for example, a subscriber is allocated a 3-Mbps CIR and a 500-KB CBS, the subscriber is guaranteed a minimum of 3 Mbps and can burst up to 500 KB of traffic and still remain within the SLA limits. If the traffic bursts above 500 KB, the traffic may be dropped or delayed.
- **PIR (Peak Information Rate)**—The PIR specifies the rate above the CIR at which traffic is allowed into the network and that may get delivered if the network is not congested. The PIR has an additional parameter associated with it called the Maximum Burst Size (MBS). The MBS is the size up to which the traffic is allowed to burst without being discarded. The MBS can be specified in KB or MB, similar to CBS. A sample service may provide a 3-Mbps CIR, 500-KB CBS, 10-Mbps PIR, and 1-MB MBS. In this case, the following behavior occurs:
  - Traffic is less than or equal to CIR (3 Mbps)—Traffic is in profile with a guaranteed delivery. Traffic is also in profile if it bursts to CBS (500 KB) and may be dropped or delayed if it bursts beyond 500 KB.
  - Traffic is more than CIR (3 Mbps) and less than PIR (10 Mbps)—Traffic is out of profile. It may get delivered if the network is not congested and the burst size is less than MBS (1 MB).
  - Traffic is more than PIR (10 Mbps)—Traffic is discarded.

Performance Parameters

The performance parameters indicate the service quality experienced by the subscriber. They consist of the following:

- Availability
- Delay
- Jitter
- Loss

Availability

Availability is specified by the following service attributes:

- **UNI Service Activation Time**—Specifies the time from when the new or modified service order is placed to the time service is activated and usable. Remember that the main value proposition that an Ethernet service claims is the ability to cut down the service activation time to hours versus months with respect to the traditional telco model.
- **UNI Mean Time to Restore (MTTR)**—Specifies the time it takes from when the UNI is unavailable to when it is restored. Unavailability can be caused by a failure such as a fiber cut.
- **EVC Service Activation Time**—Specifies the time from when a new or modified service order is placed to when the service is activated and usable. The EVC service activation time begins when all UNIs are activated. For a multipoint EVC, for example, the service is considered active when all UNIs are active and operational.
- **EVC Availability**—Specifies how often the subscriber's EVC meets or exceeds the delay, loss, and jitter service performance over the same measurement interval. If an EVC does not meet the performance criteria, it is considered unavailable.
- **EVC (MTTR)**—Specifies the time from when the EVC is unavailable to when it becomes available again. Many restoration mechanisms can be used on the physical layer (L1), the MAC layer (L2), or the network layer (L3).

Delay

Delay is a critical parameter that significantly impacts the quality of service (QoS) for real-time applications. Delay has traditionally been specified in one direction as one-way delay or end-to-end delay. The delay between two sites in the metro is an accumulation of delays, starting from one UNI at one end, going through the metro network, and going through the UNI on the other end. The delay at the UNI is affected by the line rate at the UNI connection and the supported Ethernet frame size. For example, a UNI connection with 10 Mbps and 1518-byte frame size would cause 1.2 milliseconds (ms) of transmission delay ($1518 * 8 / 10^6$).

The metro network itself introduces additional delays based on the network backbone speed and level of congestion. The delay performance is defined by the 95th percentile (95 percent)

of the delay of successfully delivered egress frames over a time interval. For example, a delay of 15 ms over 24 hours means that over a period of 24 hours, 95 percent of the "delivered" frames had a one-way delay of less than or equal to 15 ms.

The delay parameter is used in the following attributes:

- Ingress and egress bandwidth profile per CoS identifier (UNI service attribute)
- Class of service (EVC service attribute)

Jitter

Jitter is another parameter that affects the service quality. Jitter is also known as delay variation. Jitter has a very adverse effect on real-time applications such as IP telephony. The jitter parameter is used in the following service attributes:

- Ingress and egress bandwidth profile per CoS identifier (UNI service attribute)
- Class of service (EVC service attribute)

Loss

Loss indicates the percentage of Ethernet frames that are in-profile and that are not reliably delivered between UNIs over a time interval. On a P2P EVC, for example, if 100 frames have been sent from a UNI on one end and 90 frames that are in profile have been received on the other end, the loss would be (100 – 90) / 100 = 10%. Loss can have adverse effects, depending on the application. Applications such as e-mail and HTTP web browser requests can tolerate more loss than VoIP, for example. The loss parameter is used in the following attributes:

- Ingress and egress bandwidth profile per CoS identifier (UNI service attribute)
- Class of service (EVC service attribute)

Class of Service Parameters

Class of service (CoS) parameters can be defined for metro Ethernet subscribers based on various CoS identifiers, such as the following:

- **Physical port**—This is the simplest form of QoS that applies to the physical port of the UNI connection. All traffic that enters and exits the port receives the same CoS.
- **Source/destination MAC addresses**—This type of classification is used to give different types of service based on combinations of source and destination MAC addresses. While this model is very flexible, it is difficult to administer, depending on the service itself. If the customer premises equipment (CPE) at the ends of the connections are Layer 2 switches that are part of a LAN-to-LAN service, hundreds or thousands of MAC addresses might have to be monitored. On the other hand, if the CPEs are routers, the MAC addresses that are monitored are those of the router interfaces themselves. Hence, the MAC addresses are much more manageable.

- **VLAN ID**—This is a very practical way of assigning CoS if the subscriber has different services on the physical port where a service is defined by a VLAN ID (these would be the carrier-assigned VLANs).
- **802.1p value**—The 802.1p field allows the carrier to assign up to eight different levels of priorities to the customer traffic. Ethernet switches use this field to specify some basic forwarding priorities, such as that frames with priority number 7 get forwarded ahead of frames with priority number 6, and so on. This is one method that can be used to differentiate between VoIP traffic and regular traffic or between high-priority and best-effort traffic. In all practicality, service providers are unlikely to exceed two or three levels of priority, for the sake of manageability.
- **Diffserv/IP ToS**—The IP ToS field is a 3-bit field inside the IP packet that is used to provide eight different classes of service known as IP precedence. This field is similar to the 802.1p field if used for basic forwarding priorities; however, it is located inside the IP header rather than the Ethernet 802.1Q tag header. Diffserv has defined a more sophisticated CoS scheme than the simple forwarding priority scheme defined by ToS. Diffserv allows for 64 different CoS values, called Diffserv codepoints (DSCPs). Diffserv includes different per-hop behaviors (PHBs), such as Expedited Forwarding (EF) for a low delay, low-loss service, four classes of Assured Forwarding (AF) for bursty real-time and non-real-time services, Class Selector (CS) for some backward compatibility with IP ToS, and Default Forwarding (DF) for best-effort services.

  Although Diffserv gives much more flexibility to configure CoS parameters, service providers are still constrained with the issue of manageability. This is similar to the airline QoS model. Although there are so many ways to arrange seats and who sits where and so many types of food service and luggage service to offer travelers, airlines can manage at most only three or four levels of service, such as economy, economy plus, business class, and first class. Beyond that, the overhead of maintaining these services and the SLAs associated with them becomes cost-prohibitive.

Service Frame Delivery Attribute

Because the metro network behaves like a switched LAN, you must understand which frames need to flow over the network and which do not. On a typical LAN, the frames traversing the network could be data frames or control frames. Some Ethernet services support delivery of all types of Ethernet protocol data units (PDUs); others may not. To ensure the full functionality of the subscriber network, it is important to have an agreement between the subscriber and the metro carriers on which frames get carried. The EVC service attribute can define whether a particular frame is discarded, delivered unconditionally, or delivered conditionally for each ordered UNI pair. The different possibilities of the Ethernet data frames are as follows:

- **Unicast frames**—These are frames that have a specified destination MAC address. If the destination MAC address is known by the network, the frame gets delivered to the exact destination. If the MAC address is unknown, the LAN behavior is to flood the frame within the particular VLAN.

- **Multicast frames**—These are frames that are transmitted to a select group of destinations. This would be any frame with the least significant bit (LSB) of the destination address set to 1, except for broadcast, where all bits of the MAC destination address are set to 1.
- **Broadcast frames**—IEEE 802.3 defines the broadcast address as a destination MAC address, FF-FF-FF-FF-FF-FF.

Layer 2 Control Processing packets are the different L2 control-protocol packets needed for specific applications. For example, BPDU packets are needed for STP. The provider might decide to tunnel or discard these packets over the EVC, depending on the service. The following is a list of currently standardized L2 protocols that can flow over an EVC:

- **IEEE 802.3x MAC control frames**—802.3.x is an XON/XOFF flow-control mechanism that lets an Ethernet interface send a PAUSE frame in case of traffic congestion on the egress of the Ethernet switch. The 802.3x MAC control frames have destination address 01-80-C2-00-00-01.
- **Link Aggregation Control Protocol (LACP)**—This protocol allows the dynamic bundling of multiple Ethernet interfaces between two switches to form an aggregate bigger pipe. The destination MAC address for these control frames is 01-80-C2-00-00-02.
- **IEEE 802.1x port authentication**—This protocol allows a user (an Ethernet port) to be authenticated into the network via a back-end server, such as a RADIUS server. The destination MAC address is 01-80-C2-00-00-03.
- **Generic Attribute Registration Protocol (GARP)**—The destination MAC address is 01-80-C2-00-00-2X.
- **STP**—The destination MAC address is 01-80-C2-00-00-00.
- **All-bridge multicast**—The destination MAC address is 01-80-C2-00-00-10.

VLAN Tag Support Attribute

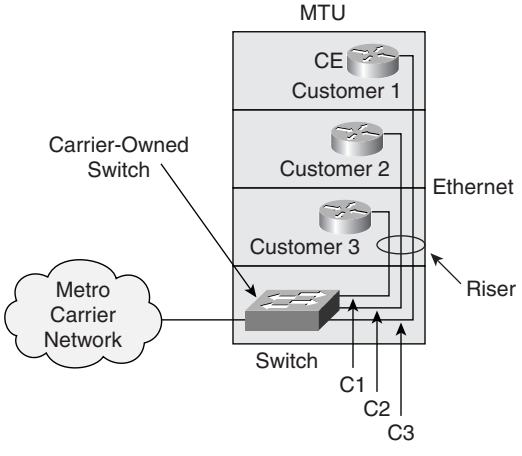
VLAN tag support provides another set of capabilities that are important for service frame delivery. Enterprise LANs are single-customer environments, meaning that the end users belong to a single organization. VLAN tags within an organization are indicative of different logical broadcast domains, such as different workgroups. Metro Ethernet creates a different environment in which the Ethernet network supports multiple enterprise networks that share the same infrastructure, and in which each enterprise network can still have its own segmentation. Support for different levels of VLANs and the ability to manipulate VLAN tags become very important.

Consider the example of an MTU building in which the metro provider installs a switch in the basement that offers multiple Ethernet connections to different small offices in the building. In this case, from a carrier perspective, each customer is identified by the physical Ethernet interface port that the customer connects to. This is shown in Figure 3-6.

Although identifying the customer itself is easy, isolating the traffic between the different customers becomes an interesting issue and requires some attention on the provider's part. Without special attention, traffic might get exchanged between the different customers in the

building through the basement switch. You have already seen in the section "L2 Switching Basics" that VLANs can be used to separate physical segments into many logical segments; however, this works in a single-customer environment, where the VLAN has a global meaning. In a multicustomer environment, each customer can have its own set of VLANs that overlap with VLANs from another customer. To work in this environment, carriers are adopting a model very similar to how Frame Relay and ATM services have been deployed. In essence, each customer is given service identifiers similar to Frame Relay data-link connection identifiers (DLCIs), which identify EVCs over which the customer's traffic travels. In the case of Ethernet, the VLAN ID given by a carrier becomes that identifier. This is illustrated in Figure 3-7.

Figure 3-6 *Ethernet in Multicustomer Environments*



In this example, the carrier needs to assign to each physical port a set of VLAN IDs that are representative of the services sold to each customer. Customer 1, for example, is assigned VLAN 10, customer 2 is assigned VLAN 20, and customer 3 is assigned VLAN 30. VLANs 10, 20, and 30 are carrier-assigned VLANs that are independent of the customer's internal VLAN assignments. To make that distinction, the MEF has given the name CE-VLANs to the customer-internal VLANs. The customers themselves can have existing VLAN assignments (CE-VLANs) that overlap with each other and the carrier's VLAN. There are two types of VLAN tag support:

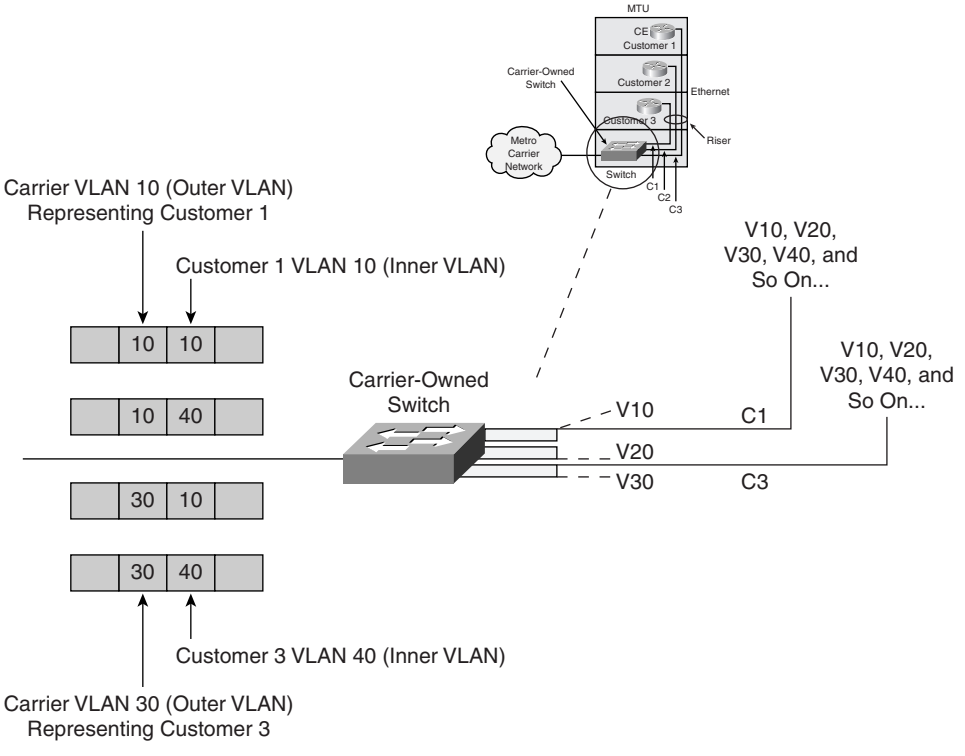
- VLAN Tag Preservation/Stacking
- VLAN Tag Translation/Swapping

VLAN Tag Preservation/Stacking

With VLAN Tag Preservation, all Ethernet frames received from the subscriber need to be carried untouched within the provider's network across the EVC. This means that the VLAN ID at the ingress of the EVC is equal to the VLAN ID on the egress. This is typical of services such as LAN extension, where the same LAN is extended between two different locations and the enterprise-internal VLAN assignments need to be preserved. Because the carrier's Ethernet

switch supports multiple customers with overlapping CE-VLANs, the carrier's switch needs to be able to stack its own VLAN assignment on top of the customer's VLAN assignment to keep the separation between the traffic of different customers. This concept is called 802.1Q-in-802.1Q or Q-in-Q stacking, as explained earlier in the section "VLAN Tagging." With Q-in-Q, the carrier VLAN ID becomes indicative of the EVC, while the customer VLAN ID (CE-VLAN) is indicative of the internals of the customer network and is hidden from the carrier's network.

Figure 3-7 *Logical Separation of Traffic and Services*



WARNING The Q-in-Q function is not standardized, and many vendors have their own variations. For the service to work, the Q-in-Q function must work on a "per-port" basis, meaning that each customer can be tagged with a different carrier VLAN tag. Some enterprise switches on the market can perform a double-tagging function; however, these switches can assign only a single VLAN-ID as a carrier ID for the whole switch. These types of switches work only if a single customer is serviced and the carrier wants to be able to carry the customer VLANs transparently within its network. These switches do not work when the carrier switch is servicing multiple customers, because it is impossible to differentiate between these customers using a single carrier VLAN tag.

VLAN Tag Translation/Swapping

VLAN Tag Translation or Swapping occurs when the VLAN tags are local to the UNI, meaning that the VLAN tag value, if it exists on one side of the EVC, is independent of the VLAN tag values on the other side. In the case where one side of the EVC supports VLAN tagging and the other side doesn't, the carrier removes the VLAN tag from the Ethernet frames before they are delivered to the destination.

Another case is two organizations that have merged and want to tie their LANs together, but the internal VLAN assignments of each organization do not match. The provider can offer a service where the VLANs are removed from one side of the EVC and are translated to the correct VLANs on the other side of the EVC. Without this service, the only way to join the two organizations is via IP routing, which ignores the VLAN assignments and delivers the traffic based on IP addresses.

Another example of tag translation is a scenario where different customers are given Internet connectivity to an ISP. The carrier gives each customer a separate EVC. The carrier assigns its own VLAN-ID to the EVC and strips the VLAN tag before handing off the traffic to the ISP. This is illustrated in Figure 3-8.

Figure 3-8 *VLAN Translation*

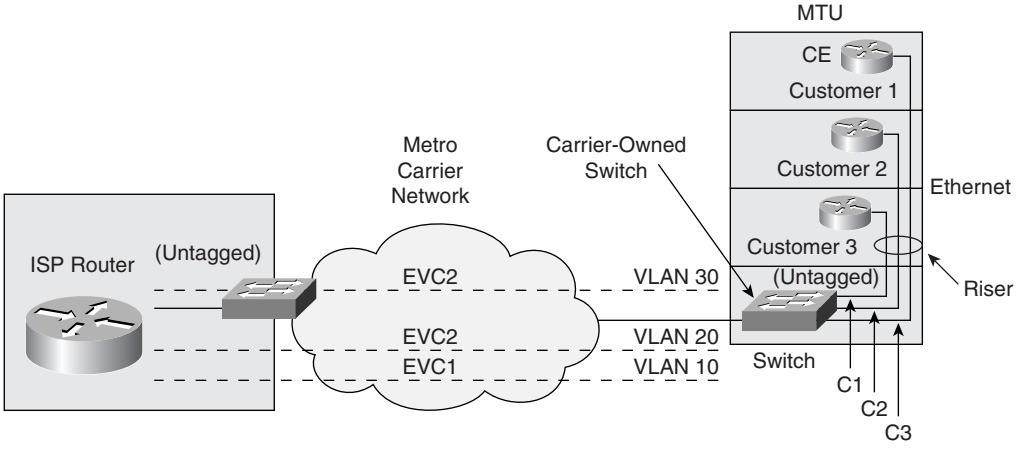


Figure 3-8 shows the metro carrier delivering Internet connectivity to three customers. The carrier is receiving untagged frames from the CPE routers located at each customer premises. The carrier inserts a VLAN tag 10 for all of customer 1's traffic, VLAN 20 for customer 2's traffic, and VLAN 30 for customer 3's traffic. The carrier uses the VLAN tags to separate the three customers' traffic within its own network. At the point of presence (POP), the VLAN tags are removed from all EVCs and handed off to an ISP router, which is offering the Internet IP service.

Service Multiplexing Attribute

Service multiplexing is used to support multiple instances of EVCs on the same physical connection. This allows the same customer to have different services with the same Ethernet wire.

Bundling Attribute

The Bundling service attribute enables two or more VLAN IDs to be mapped to a single EVC at a UNI. With bundling, the provider and subscriber must agree on the VLAN IDs used at the UNI and the mapping between each VLAN ID and a specific EVC. A special case of bundling is where every VLAN ID at the UNI interface maps to a single EVC. This service attribute is called *all-to-one bundling.*

Security Filters Attribute

Security filters are MAC access lists that the carrier uses to block certain addresses from flowing over the EVC. This could be an additional service the carrier can offer at the request of the subscriber who would like a level of protection against certain MAC addresses. MAC addresses that match a certain access list could be dropped or allowed.

Tables 3-1 and 3-2 summarize the Ethernet service attributes and their associated parameters for UNI and EVCs.

Table 3-1 *UNI Service Attributes*

UNI Service Attribute	Parameter Values or Range of Values
Physical medium	A standard Ethernet physical interface.
Speed	10 Mbps, 100 Mbps, 1 Gbps, or 10 Gbps.
Mode	Full-duplex or autospeed negotiation.
MAC layer	Ethernet and/or IEEE 802.3-2002.
Service multiplexing	Yes or no. If yes, all-to-one bundling must be no.
Bundling	Yes or no. Must be no if all-to-one bundling is yes and yes if all-to-one bundling is no.
All-to-one bundling	Yes or no. If yes, service multiplexing and bundling must be no. Must be no if bundling is yes.
Ingress and egress bandwidth profile per UNI	No or one of the following parameters: CIR, CBS, PIR, MBS. If no, no bandwidth profile per UNI is set; otherwise, the traffic parameters CIR, CBS, PIR, and MBS need to be set.

continues

Table 3-1 *UNI Service Attributes (Continued)*

UNI Service Attribute	Parameter Values or Range of Values
Ingress and egress bandwidth profile per EVC	No or one of the following parameters: CIR, CBS, PIR, MBS.
Ingress and egress bandwidth profile per CoS identifier	No or one of the following parameters: CIR, CBS, PIR, MBS. If one of the parameters is chosen, specify the CoS identifier, Delay value, Jitter value, Loss value. If no, no bandwidth profile per CoS identifier is set; otherwise, the traffic parameters CIR, CBS, PIR, and MBS need to be set.
Ingress and egress bandwidth profile per destination UNI per EVC	No or one of the following parameters: CIR, CBS, PIR, MBS.
Egress bandwidth profile per source UNI per EVC	No or one of the following parameters: CIR, CBS, PIR, MBS.
Layer 2 Control Protocol processing	Process, discard, or pass to EVC the following control protocol frames: • IEEE 802.3x MAC control • Link Aggregation Control Protocol (LACP) • IEEE 802.1x port authentication • Generic Attribute Registration Protocol (GARP) • STP • Protocols multicast to all bridges in a bridged LAN
UNI service activation time	Time value

Table 3-2 *EVC Service Attributes*

EVC Service Attribute	Type of Parameter Value
EVC Type	P2P or MP2MP
CE-VLAN ID preservation	Yes or no
CE-VLAN CoS preservation	Yes or no
Unicast frame delivery	Discard, deliver unconditionally, or deliver conditionally for each ordered UNI pair
Multicast frame delivery	Discard, deliver unconditionally, or deliver conditionally for each ordered UNI pair
Broadcast frame delivery	Discard, deliver unconditionally, or deliver conditionally for each ordered UNI pair

Table 3-2 *EVC Service Attributes (Continued)*

EVC Service Attribute	Type of Parameter Value
Layer 2 Control Protocol processing	Discard or tunnel the following control frames: • IEEE 802.3x MAC control • Link Aggregation Control Protocol (LACP) • IEEE 802.1x port authentication • Generic Attribute Registration Protocol (GARP) • STP • Protocols multicast to all bridges in a bridged LAN
EVC service activation time	Time value
EVC availability	Time value
EVC mean time to restore	Time value
Class of service	CoS identifier, Delay value, Jitter value, Loss value This assigns the Class of Service Identifier to the EVC

Example of an L2 Metro Ethernet Service

This section gives an example of an L2 metro Ethernet service and how all the parameters defined by the MEF are applied. The example attempts to highlight many of the definitions and concepts discussed in this chapter.

If you have noticed, the concept of VPNs is inherent in L2 Ethernet switching. The carrier VLAN is actually a VPN, and all customer sites within the same carrier VLAN form their own user group and exchange traffic independent of other customers on separate VLANs.

The issue of security arises in dealing with VLAN isolation between customers; however, because the metro network is owned by a central entity (such as the metro carrier), security is enforced. First of all, the access switches in the customer basement are owned and administered by the carrier, so physical access is prevented. Second, the VLANs that are switched in the network are assigned by the carrier, so VLAN isolation is guaranteed. Of course, misconfiguration of switches and VLAN IDs could cause traffic to be mixed, but this problem can occur with any technology used, not just Ethernet. Issues of security always arise in public networks whether they are Ethernet, IP, MPLS, or Frame Relay networks. The only definite measure to ensure security is to have the customer-to-customer traffic encrypted at the customer sites and to have the customers administer that encryption.

Figure 3-9 shows an example of an L2 metro Ethernet VPN. This example attempts to show in a practical way how many of the parameters and the concepts that are discussed in this chapter are used.

Figure 3-9 *All-Ethernet L2 Metro Service Example*

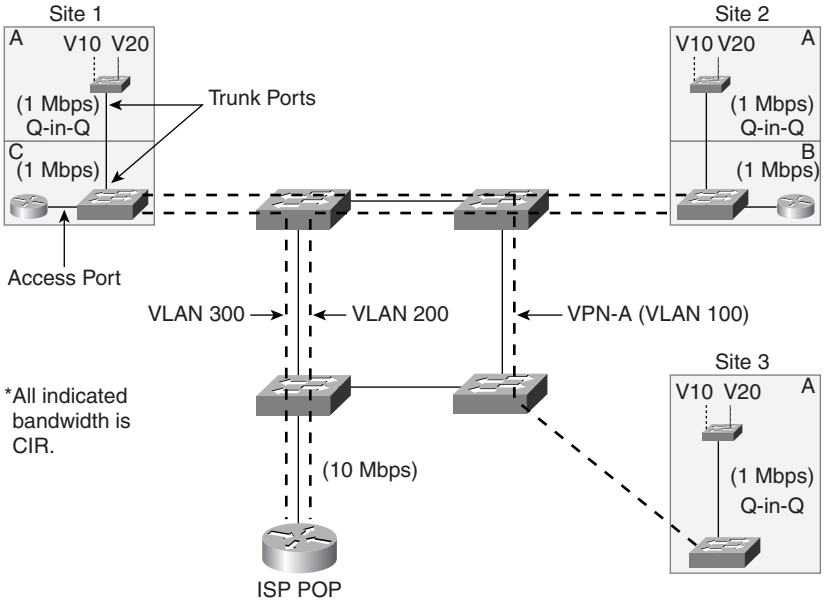


Figure 3-9 shows a metro carrier offering an L2 MP2MP VPN service to customer A and a packet leased-line service (comparable to a traditional T1 leased line) to an ISP. In turn, the ISP is offering Internet service to customers B and C. It is assumed that customer A connects to the carrier via L2 Ethernet switches and customers B and C connect via IP routers. Notice the difference between access ports and trunk ports on the Ethernet switches. The ports connecting the customer's Ethernet switch to the carrier's Ethernet switch are trunk ports, because these ports are carrying multiple VLANs between the two switches. When the carrier's switch port is configured for Q-in-Q, it encapsulates the customers' CE-VLAN tags VLAN 10 and VLAN 20 inside the carrier VLAN 100. On the other hand, the ports connecting the customer router with the carrier switch are access ports and are carrying untagged traffic from the router. Tables 3-3 and 3-4 describe the UNI and EVC service attributes for customers A, B, and C as defined by the MEF.

Table 3-3 *Customer A E-LAN UNI Service Attributes*

Customer A E-LAN UNI Service Attribute	**Parameter Values or Range of Values**
Physical medium	Standard Ethernet physical interfaces
Speed	100 Mbps site 1, 10 Mbps sites 2 and 3
Mode	Full duplex all sites
MAC layer	IEEE 802.3-2002
Service multiplexing	No

Table 3-3 *Customer A E-LAN UNI Service Attributes (Continued)*

Customer A E-LAN UNI Service Attribute	Parameter Values or Range of Values
Bundling	No
All-to-one bundling	Yes
Ingress and egress bandwidth profile per CoS identifier	All sites CoS 1: • CIR = 1 Mbps, CBS = 100 KB, PIR = 2 Mbps, MBS = 100 KB • CoS ID = 802.1p 6–7 • Delay < 10 ms, Loss < 1% All sites CoS 0: • CIR = 1 Mbps, CBS = 100 KB, PIR = 10 Mbps, MBS = 100 KB • CoS ID = 802.1p 0–5, Delay < 35 ms, Loss < 2%
Layer 2 Control Protocol processing	• Process IEEE 802.3x MAC control • Process Link Aggregation Control Protocol (LACP) • Process IEEE 802.1x port authentication • Pass Generic Attribute Registration Protocol (GARP) • Pass STP • Pass protocols multicast to all bridges in a bridged LAN
UNI service activation time	One hour after equipment is installed

Note in Table 3-3 that customer A is given only one MP2P EVC; hence, there is no service multiplexing. All customer VLANs 10 and 20 are mapped to the MP2MP EVC in the form of carrier VLAN 100. Customer A is given two Class of Service profiles—CoS 1 and CoS 0. Each profile has its set of performance attributes. Profile 1, for example, is applied to high-priority traffic, as indicated by 802.1p priority levels 6 and 7. Profile 0 is lower priority, with less-stringent performance parameters. For customer A, the metro carrier processes the 802.3x and LACP frames on the UNI connection and passes other L2 control traffic that belongs to the customer. Passing the STP control packets, for example, prevents any potential loops within the customer network, in case the customer has any L2 backdoor direct connection between its different sites.

Table 3-4 *Customer A E-LAN EVC Service Attributes*

Customer A E-LAN EVC Service Attribute	Type of Parameter Value
EVC type	MP2MP
CE-VLAN ID preservation	Yes

continues

Table 3-4 *Customer A E-LAN EVC Service Attributes (Continued)*

Customer A E-LAN EVC Service Attribute	Type of Parameter Value
CE-VLAN CoS preservation	Yes
Unicast frame delivery	Deliver unconditionally for each UNI pair
Multicast frame delivery	Deliver unconditionally for each UNI pair
Broadcast frame delivery	Deliver unconditionally for each UNI pair
Layer 2 Control Protocol processing	Tunnel the following control frames: • IEEE 802.3x MAC control • Link Aggregation Control Protocol (LACP) • IEEE 802.1x port authentication • Generic Attribute Registration Protocol (GARP) • STP • Protocols multicast to all bridges in a bridged LAN
EVC service activation time	Twenty minutes after UNI is operational
EVC availability	Three hours
EVC mean time to restore	One hour
Class of service	All sites CoS 1: • CoS ID = 802.1p 6–7 • Delay < 10 ms, Loss < 1%, Jitter (value) All sites CoS 0: • CoS ID = 802.1p 0–5, Delay < 35 ms, Loss < 2%, Jitter (value)

The EVC service parameters for customer A indicate that the EVC is an MP2MP connection and the carrier transparently moves the customer VLANs between sites. The carrier does this using Q-in-Q tag stacking with a carrier VLAN ID of 100. The carrier also makes sure that the 802.1p priority fields that the customer sends are still carried within the network. Note that the carrier allocates priority within its network whichever way it wants as long as the carrier delivers the SLA agreed upon with the customer as described in the CoS profiles. For customer A, the carrier passes all unicast, multicast, and broadcast traffic and also tunnels all L2 protocols between the different sites.

Tables 3-5 and 3-6 describe customers B and C and ISP POP service profile for the Internet connectivity service. These are the service attributes and associated parameters for customers

B and C as well as the service attributes and associated parameters for the ISP POP offering Internet connectivity to these customers.

Table 3-5 *Customers B and C and ISP POP UNI Service Attributes*

Customers B and C and ISP POP Internet Access UNI Service Attribute	**Parameter Values or Range of Values**
Physical medium	Standard Ethernet physical interfaces
Speed	10 Mbps for customers B and C, 100 Mbps for the ISP POP
Mode	Full duplex all sites
MAC layer	IEEE 802.3-2002
Service multiplexing	Yes, only at ISP POP UNI
Bundling	No
All-to-one bundling	No
Ingress and egress bandwidth profile per EVC	Customers B and C CIR = 1 Mbps, CBS = 100 KB, PIR = 2 Mbps, MBS = 100 KB ISP POP CIR = 10 Mbps, CBS = 1 MB, PIR = 100 Mbps, MBS = 1 MB
Layer 2 Control Protocol processing	Discard the following control frames: • IEEE 802.3x MAC control • Link Aggregation Control Protocol (LACP) • IEEE 802.1x port authentication • Generic Attribute Registration Protocol (GARP) • STP • Protocols multicast to all bridges in a bridged LAN
UNI service activation time	One hour after equipment is installed

For customers B and C and ISP POP UNI service parameters, because two different P2P EVCs (carrier VLANs 200 and 300) are configured between the customers and the ISP POP, service multiplexing occurs at the ISP UNI connection where two EVCs are multiplexed on the same physical connection. For this Internet access scenario, routers are the customer premises equipment, so it is unlikely that the customer will send any L2 control-protocol packets to the carrier. In any case, all L2 control-protocol packets are discarded if any occur.

Table 3-6 *Customers B and C and ISP POP EVC Service Attributes*

Customers B and C and ISP POP Internet Access EVC Service Attribute	Type of Parameter Value
EVC type	P2P
CE-VLAN ID preservation	No; mapped VLAN ID for provider use
CE-VLAN CoS preservation	No
Unicast frame delivery	Deliver unconditionally for each UNI pair
Multicast frame delivery	Deliver unconditionally for each UNI pair
Broadcast frame delivery	Deliver unconditionally for each UNI pair
Layer 2 Control Protocol processing	Discard the following control frames: • IEEE 802.3x MAC control • Link Aggregation Control Protocol (LACP) • IEEE 802.1x port authentication • Generic Attribute Registration Protocol (GARP) • STP • Protocols multicast to all bridges in a bridged LAN
EVC service activation time	Twenty minutes after UNI is operational
EVC availability	Three hours
EVC mean time to restore	One hour
Class of service	One CoS service is supported: Delay < 30 ms, Loss < 1%, Jitter (value)

The EVC parameters indicate that the carrier is not preserving any customer VLANs or CoS info. Also, because this is an Internet access service, normally the provider receives untagged frames from the CPE router. The provider can map those frames to carrier VLANs 200 and 300 if it needs to separate the traffic in its network. The VLAN IDs are normally stripped off before given to the ISP router.

Challenges with All-Ethernet Metro Networks

All-Ethernet metro networks pose many scalability and reliability challenges. The following are some of the issues that arise with an all-Ethernet control plane:

- Restrictions on the number of customers
- Service monitoring
- Scaling the L2 backbone

- Service provisioning
- Interworking with legacy deployments

The following sections describe each of these challenges.

Restrictions on the Number of Customers

The Ethernet control plane restricts the carrier to 4096 customers, because the 802.1Q defines 12 bits that can be used as a VLAN ID, which restricts the number of VLANs to $2^{12} = 4096$. Remember that although Q-in-Q allows the customer VLANs (CE-VLANs) to be hidden from the carrier network, the carrier is still restricted to 4096 VLAN IDs that are global within its network. For many operators that are experimenting with the metro Ethernet service, the 4096 number seems good enough for an experimental network but presents a long-term roadblock if the service is to grow substantially.

Service Monitoring

Ethernet does not have an embedded mechanism that lends to service monitoring. With Frame Relay LMI, for example, service monitoring and service integrity are facilitated via messages that report the status of the PVC. Ethernet service monitoring requires additional control-plane intelligence. New Link Management Interface (LMI) protocols need to be defined and instituted between the service provider network and the CPE to allow the customer to discover the different EVCs that exist on the UNI connection. The LMI could learn the CE-VLAN to EVC map and could learn the different service parameters such as bandwidth profiles. Other protocols need to be defined to discover the integrity of the EVC in case of possible failures. You have seen in the previous section how performance parameters could indicate the availability of an EVC. Protocols to extract information from the UNI and EVC are needed to make such information usable.

Scaling the L2 Backbone

A metro carrier that is building an all-Ethernet network is at the mercy of STP. STP blocks Ethernet ports to prevent network loops. Traffic engineering (discussed in Chapter 5, “MPLS Traffic Engineering”) is normally a major requirement for carriers to have control over network bandwidth and traffic trajectory. It would seem very odd for any carrier to have the traffic flow in its network be dependant on loop prevention rather than true bandwidth-optimization metrics.

Service Provisioning

Carrying a VLAN through the network is not a simple task. Any time a new carrier VLAN is created (a new VPN), care must be taken to configure that VLAN across all switches that need to participate in that VPN. The lack of any signaling protocols that allow VPN information to

be exchanged makes the task manual and tedious. Early adopters of metro Ethernet have endured the pains of carrying VLANs across many switches. Even with the adoption of new protocols such as 802.1s ("Amendment to 802.1Q (TM) Virtual Bridged Local Area Networks: Multiple Spanning Trees"), the task of scaling the network is almost impossible.

Interworking with Legacy Deployments

Another challenge facing Ethernet deployments is interworking with legacy deployments such as existing Frame Relay and ATM networks. Frame Relay has been widely deployed by many enterprises as a WAN service. Remote offices are connected to headquarters via P2P Frame Relay circuits forming a hub-and-spoke topology. Enterprises that want to adopt Ethernet as an access technology expect the carrier to provide a means to connect the new sites enabled with Ethernet access with existing headquarters sites already enabled with Frame Relay. This means that a function must exist in the network that enables Frame Relay and Ethernet services to work together.

The IETF has standardized in RFC 2427, *Multiprotocol Interconnect over Frame Relay,* how to carry different protocols over Frame Relay, including Ethernet. In some other cases, the Ethernet and Frame Relay access networks are connected by an ATM core network. In this case, two service-interworking functions need to happen, one between Ethernet and ATM and another between ATM and Frame Relay. Ethernet-to-ATM interworking is achieved using RFC 2684, and ATM-to-Frame Relay interworking is achieved via the Frame Relay Forum specification FRF 8.1. Figure 3-10 illustrates the service-interworking functions.

Figure 3-10 *Service Interworking*

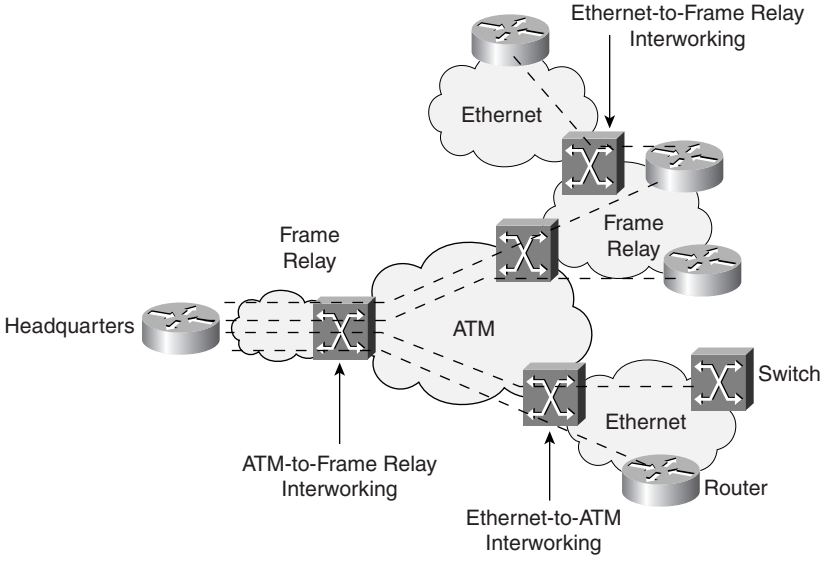


Figure 3-10 shows a scenario in which an enterprise headquarters is connected to its remote sites via Frame Relay connections carried over an ATM network. The different service-interworking functions are displayed to allow such networks to operate. For service interworking, two encapsulation methods are defined: one is bridged, and the other is routed. Both sides of the connection are either bridged or routed. Some challenges might exist if one end of the connection is connected to a LAN switch, and hence bridged, while the other end is connected to a router. Other issues will arise because of the different Address Resolution Protocol (ARP) formats between the different technologies, such as Ethernet, Frame Relay, and ATM. Some vendors are attempting to solve these problems with special software enhancements; however, such practices are still experimental and evolving.

It is all these challenges that motivated the emergence of hybrid architectures consisting of multiple L2 domains that are connected via L3 IP/MPLS cores. The network can scale because L2 Ethernet would be constrained to more-controlled access deployments that limit the VLAN and STP inefficiencies. The network can then be scaled by building a reliable IP/MPLS core. This is discussed in Chapter 4, "Hybrid L2 and L3 IP/MPLS Networks."

Conclusion

This chapter has discussed many aspects of metro Ethernet services. The MEF is active in defining the characteristics of these services, including the service definitions and framework and the many service attributes that make up the services. Defining the right traffic and performance parameters, class of service, service frame delivery, and other aspects ensures that buyers and users of the service understand what they are paying for and also helps service providers communicate their capabilities.