



Introduction to Cisco MPLS VPN Technology

Technology Overview

The Cisco VPN Solutions Center: MPLS Solution, a modular suite of network and service management applications, is a network management system that defines and monitors virtual private network (VPN) services for service providers. MPLS VPN Solution allows service providers to provision and manage intranet and extranet VPNs. The product provides the aspect of operations management that addresses flow-through provisioning, service auditing, and Service Level Agreement (SLA) measurement of IP-based MPLS VPN environments. Multiprotocol Label Switching (MPLS) is an emerging industry standard upon which tag switching is based.

MPLS VPN Solution is a scalable, provider-focused VPN technology that allows service providers to plan, provision, and manage for IP VPN services according to a customer's service level agreement. This product complements Cisco's MPLS-based VPN solutions by simplifying the provisioning, service assurance, and billing processes, thereby reducing the cost of deploying and operating VPN services. MPLS VPN Solution does not contain a billing application, but the product enables billing by providing the usage data on services that a billing engine can process.

MPLS VPN Solution focuses on provisioning, auditing, and monitoring the links between the customer's routers through the provider's network. This product deals only with the provider's edge routers and the customer's edge routers. A customer edge router (CE) is connected to a provider edge router (PE) in such a way that the customer's traffic is encapsulated and transparently sent to other CEs, thus creating a virtual private network. CEs advertise routes to the VPN for all the devices in their site. The MPLS VPN Solution provisioning engine accesses the configuration files on both the CE and PE to compute the necessary changes to those files that are required to support the service on the PE-CE link.

Using the MPLS VPN Solution software, service providers can do the following:

- Provision IP-based MPLS VPN services
- Generate audit reports for service requests
- Perform data collection to measure SLA performance
- Evaluate service usage for each VPN

An MPLS VPN consists of a set of sites that are interconnected by means of an MPLS provider core network. At each site, there are one or more CEs, which attach to one or more PEs. PEs use the Border Gateway Protocol-Multiprotocol (MP-BGP) to dynamically communicate with each other.

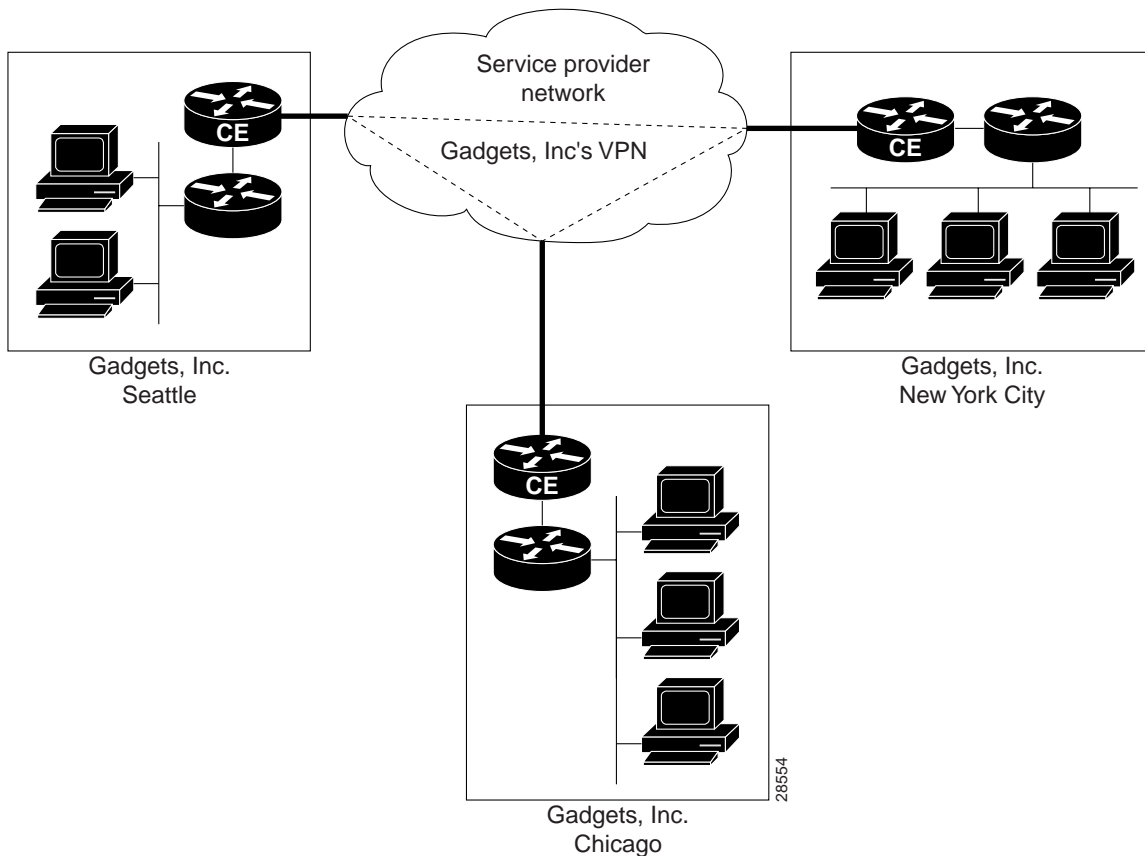
It is not required that the set of IPv4 addresses used in any two VPNs be mutually exclusive because the PEs translate IPv4 addresses into IPv4 VPN entities by using MP-BGP with extended community attributes.

The set of IP addresses used in a VPN, however, must be exclusive of the set of addresses used in the provider network. Every CE must be able to address the PEs to which it is directly attached. Thus, the IP addresses of the PEs must not be duplicated in any VPN.

The Customer's and Provider's View of the Network

From the customer's point of view, they see their internal routers communicating with their customer edge routers (CEs) from one site to another through a VPN managed by the service provider (see Figure 1-1).

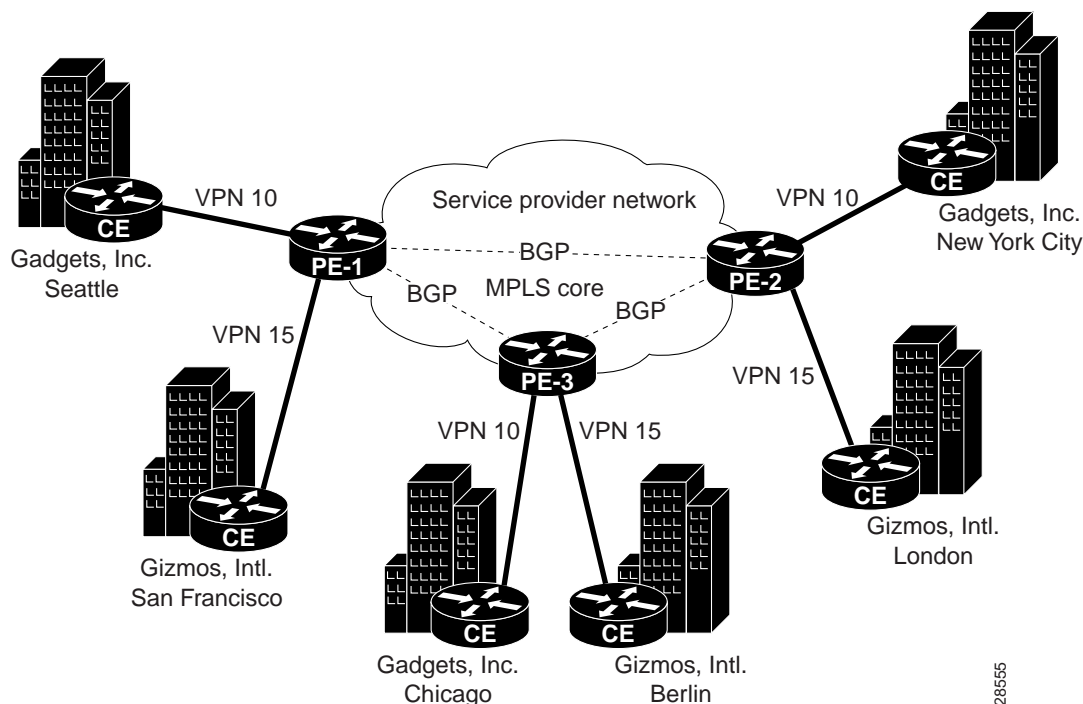
Figure 1-1 The Customer's View of the Network



This simple view of the customer's network is the advantage of employing VPNs: the customer experiences direct communication to their sites as though they had their own private network, even though their traffic is traversing a public network infrastructure and they are sharing that infrastructure with other businesses.

The service provider's view of the network is naturally very different, as shown in Figure 1-2. This illustration shows two different customers, with each customer having a single VPN. A customer can, however, have multiple VPNs.

Figure 1-2 Service Provider's View of the Network



28555

About PEs

At the edge of the provider network are provider edge routers (PEs). Within the provider network are other provider routers as needed (often designated as P routers) that communicate with each other and the PEs via the Border Gateway Protocol-Multiprotocol (MP-BGP). Note that in this model, the service provider need only provision the links between the PEs and CEs.

PEs maintain separate routing tables called VPN routing and forwarding tables (VRFs). The VRFs contain the routes for directly connected VPN sites only. (For more information about VRFs, see the “VPN Routing and Forwarding Tables (VRFs)” section on page 1-7). PEs exchange VPN-IPv4 updates through MP-iBGP sessions. These updates contain VPN-IPv4 addresses and labels. The PE originating the route is the next hop of the route. PE addresses are referred to as host routes into the core interior gateway protocol.

Benefits

MPLS-based VPNs provide the following benefits:

- A platform for rapid deployment of additional value-added IP services, including intranets, extranets, voice, multimedia, and network commerce
- Privacy and security equal to Layer-2 VPNs by constraining the distribution of a VPN's routes to only those routers that are members of that VPN, and by using MPLS for forwarding
- Seamless integration with customer intranets
- Increased scalability with thousands of sites per VPN and hundreds of thousands of VPNs per service provider

- IP Class of Service (CoS) with support for multiple classes of service within a VPN, as well as priorities among VPNs
- Easy management of VPN membership and rapid deployment of new VPNs
- Scalable any-to-any connectivity for extended intranets and extranets that encompass multiple businesses

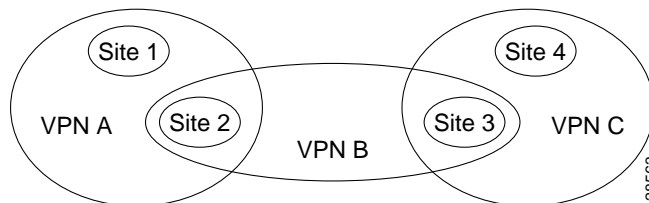
About MPLS VPNs

A virtual private network (VPN) is a network in which customer connectivity to multiple sites is deployed on a shared infrastructure with the same administrative policies as a private network. The path between two systems in a VPN, and the characteristics of that path, may also be determined (wholly or partially) by policy. Whether a system in a particular VPN is allowed to communicate with systems not in the same VPN is also a matter of policy.

In MPLS VPN, a VPN generally consists of a set of sites that are interconnected by means of an MPLS provider core network, but it is also possible to apply different policies to different systems that are located at the same site. Policies can also be applied to systems that dial in; the chosen policies would be based on the dial-in authentication processes.

A given set of systems can be in one or more VPNs. A VPN can consist of sites (or systems) that are all from the same enterprise (intranet), or from different enterprises (extranet); it may consist of sites (or systems) that all attach to the same service provider backbone, or to different service provider backbones.

Figure 1-3 VPNs Sharing Sites



MPLS-based VPNs are created in Layer 3 and are based on the peer model, which makes them more scalable and easier to build and manage than conventional VPNs. In addition, value-added services, such as application and data hosting, network commerce, and telephony services, can easily be targeted and deployed to a particular MPLS VPN because the service provider backbone recognizes each MPLS VPN as a secure, connectionless IP network.

The MPLS VPN model is a true peer VPN model that enforces traffic separations by assigning unique VPN route forwarding tables (VRFs) to each customer's VPN. Thus, users in a specific VPN cannot see traffic outside their VPN. Traffic separation occurs without tunneling or encryption because it is built directly into the network. (For more information on VRFs, see the "VPN Routing and Forwarding Tables (VRFs)" section on page 1-7.

The service provider's backbone is comprised of the PE and its provider routers. MPLS VPN provides the ability that the routing information about a particular VPN be present *only* in those PE routers that attach to that VPN.

Characteristics of MPLS VPNs

MPLS VPNs have the following characteristics:

- Multiprotocol Border Gateway Protocol-Multiprotocol (MP-BGP) extensions are used to encode customer IPv4 address prefixes into unique VPN-IPv4 Network Layer Reachability Information (NLRI) values.

NLRI refers to a destination address in MP-BGP, so NLRI is considered “one routing unit.” In the context of IPv4 MP-BGP, NLRI refers to a network prefix/prefix length pair that is carried in the BGP4 routing updates.

- Extended MP-BGP community attributes are used to control the distribution of customer routes.
- Each customer route is associated with an MPLS label, which is assigned by the provider edge router that originates the route. The label is then employed to direct data packets to the correct egress customer edge router.

When a data packet is forwarded across the provider backbone, two labels are used. The first label directs the packet to the appropriate egress PE; the second label indicates how that egress PE should forward the packet.

- Cisco MPLS CoS and QoS mechanisms provide service differentiation among customer data packets. For more information, see the “Quality of Service and Class of Service” section on page 1-13.
- The link between the PE and CE routers uses standard IP forwarding.

The PE associates each CE with a per-site forwarding table that contains only the set of routes available to that CE.

Principal Technologies

There are four principal technologies that make it possible to build MPLS-based VPNs:

- Multiprotocol Border Gateway Protocol-Multiprotocol (MP-BGP) between PEs carries CE routing information
- Route filtering based on the VPN route target extended MP-BGP community attribute
- MPLS forwarding carries packets between PEs (across the service provider backbone)
- Each PE has multiple VPN routing and forwarding instances (VRFs)

Intranets and Extranets

If all the sites in a VPN are owned by the same enterprise, the VPN is a corporate *intranet*. If the various sites in a VPN are owned by different enterprises, the VPN is an *extranet*. A site can be in more than one VPN. Both intranets and extranets are regarded as VPNs.

While the basic unit of interconnection is the site, the MPLS VPN architecture allows a finer degree of granularity in the control of interconnectivity. For example, at a given site, it may be desirable to allow only certain specified systems to connect to certain other sites. That is, certain systems at a site may be members of an intranet as well as members of one or more extranets, while other systems at the same site may be restricted to being members of the intranet only.

A CE router can be in multiple VPNs, although it can only be in a single site. When a CE router is in multiple VPNs, one of these VPNs is considered its primary VPN. In general, a CE router’s primary VPN is the intranet that includes the CE router’s site. A PE router may attach to CE routers in any number of different sites, whether those CE routers are in the same or in different VPNs. A CE router may, for robustness, attach to multiple PE routers. A PE router attaches to a particular VPN if it is a router adjacent to a CE router that is in that VPN.

Ensuring Security

Security in MPLS VPN Solution networks is delivered through a combination of MP-BGP and IP address resolution. In addition, service providers can ensure that VPNs are isolated from each other.

MP-BGP Security Features

Multiprotocol BGP is a routing information distribution protocol that, through employing multiprotocol extensions and community attributes, defines who can talk to whom. VPN membership depends upon logical ports entering the VPN, where MP-BGP assigns a unique Route Distinguisher (RD) value (see “Route Distinguishers and Route Targets” below).

RDs are unknown to end users, making it impossible to enter the network on another access port and spoof a flow. Only preassigned ports are allowed to participate in the VPN. In an MPLS VPN, MP-BGP distributes forwarding information base (FIB) tables about VPNs to members of the same VPN only, providing native security via logical VPN traffic separation. Furthermore, IBGP PE routing peers can perform TCP segment protection using the MD5 Signature Option when establishing IBGP peering relationships, further reducing the likelihood of introducing spoofed TCP segments into the IBGP connection stream among PE routers (for information on the MD5 Signature Option, see RFC 2385).

The service provider, not the customer, associates a specific VPN with each interface when provisioning the VPN. Users can only participate in an intranet or extranet if they reside on the correct physical or logical port and have the proper RD. This setup makes a Cisco MPLS VPN virtually impossible to enter.

Within the core, a standard Interior Gateway Protocol (IGP) such as OSPF or IS-IS distributes routing information. Provider edge routers set up paths among one another using LDP to communicate label-binding information. Label binding information for external (customer) routes is distributed among PE routers using MP-BGP multiprotocol extensions instead of LDP, because they easily attach to VPN IP information already being distributed.

The MP-BGP community attribute constrains the scope of reachability information. MP-BGP maps FIB tables to provider edge routers belonging to only a particular VPN, instead of updating all edge routers in the service provider network.

Security Through IP Address Resolution

MPLS VPN networks are easier to integrate with IP-based customer networks. Subscribers can seamlessly interconnect with a provider service without changing their intranet applications because MPLS-based networks have built-in application awareness. Customers can even transparently use their existing IP address space without Network Address Translator (NAT) because each VPN has a unique identifier.

MPLS VPNs remain unaware of one another. Traffic is separated among VPNs using a logically distinct forwarding table and RD for each VPN. Based on the incoming interface, the PE selects a specific forwarding table, which lists only valid destinations in the VPN. To create extranets, a provider explicitly configures reachability among VPNs.

The forwarding table for a PE contains only address entries for members of the same VPN. The PE rejects requests for addresses not listed in its forwarding table. By implementing a logically separate forwarding table for each VPN, each VPN itself becomes a private, connectionless network built on a shared infrastructure.

IP limits the size of an address to 32 bits in the packet header. The VPN IP address adds 64 bits in front of the header, creating an extended address in routing tables that classical IP cannot forward. MPLS solves this problem by forwarding traffic based on labels, so one can use MPLS to bind VPN IP routes to label-switched paths. PEs are concerned with reading labels, not packet headers. MPLS manages forwarding through the provider's MPLS core. Since labels only exist for valid destinations, this is how MPLS delivers both security and scalability.

When a virtual circuit is provided using the overlay model, the egress interface for any particular data packet is a function solely of the packet's ingress interface; the IP destination address of the packet does not determine its path in the backbone network. Thus, unauthorized communication into or out of a VPN is prevented.

In MPLS VPNs, a packet received by the backbone is first associated with a particular VPN, by stipulating that all packets received on a certain interface (or subinterface) belong to a certain VPN. Then its IP address is looked up in the forwarding table associated with that VPN. The routes in that forwarding table are specific to the VPN of the received packet.

In this way, the ingress interface determines a set of possible egress interfaces, and the packet's IP destination address is used to choose from among that set. This prevents unauthorized communication into and out of a VPN.

Ensuring VPN Isolation

To maintain proper isolation of one VPN from another, it is important that the provider routers not accept a labeled packet from any adjacent PE unless the following conditions are met:

- The label at the top of the label stack was actually distributed by the provider router to the PE device.
- The provider router can determine that use of that label will cause the packet to exit the backbone before any labels lower in the stack and the IP header will be inspected.

These restrictions are necessary to prevent packets from entering a VPN where they do not belong.

The VRF tables in a PE are used only for packets arriving from a CE that is directly attached to the PE device. They are not used for routing packets arriving from other routers that belong to the service provider backbone. As a result, there may be multiple different routes to the same system, where the route followed by a given packet is determined by the site from which the packet enters the backbone. So one may have one route to a given IP network for packets from the extranet (where the route leads to a firewall), and a different route to the same network for packets from the intranet.

VPN Routing and Forwarding Tables (VRFs)

The VPN routing and forwarding table (VRF) is a key element in the MPLS VPN technology. VRFs exist on PEs only. A VRF is a routing table instance, and more than one VRF can exist on a PE. A VPN can contain one or more VRFs on a PE. The VRF contains routes that should be available to a particular set of sites. VRFs use CEF technology, therefore the VPN must be CEF-enabled.

A VRF is associated with the following elements:

- IP routing table
- Derived forwarding table, based on the Cisco Express Forwarding (CEF) technology
- A set of interfaces that use the derived forwarding table
- A set of routing protocols and routing peers that inject information into the VRF

Each PE maintains one or more VRFs. MPLS VPN software looks up a particular packet's IP destination address in the appropriate VRF only if that packet arrived directly through an interface that is associated with that VRF. The so-called "color" MPLS label tells the destination PE to check the VRF for the appropriate VPN so that it can deliver the packet to the correct CE and finally to the local host machine.

A VRF is named based on the VPN or VPNs it services, and on the role of the CE in the topology. The schemes for the VRF names are as follows:

The VRF name for a hub: `ip vrf vx[VPN_name]`

The *x* parameter is a number assigned to make the VRF name unique.

For example, if we consider a VPN called Blue, then a VRF for a hub CE would be called:

```
ip vrf V1:blue
```

A VRF for a spoke CE in the Blue VPN would be called:

```
ip vrf V1:blue-s
```

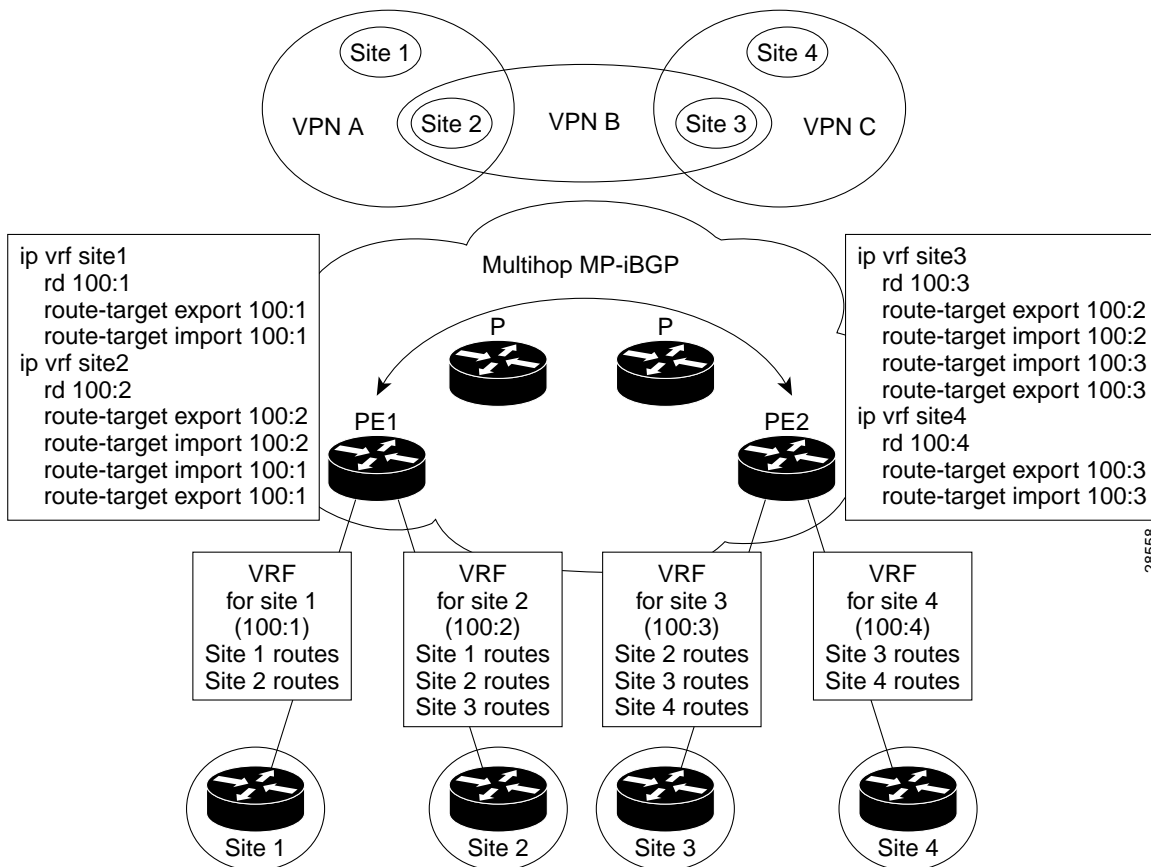
A VRF for an extranet VPN topology in the Green VPN would be called:

```
ip vrf V1:green-etc
```

Thus, you can read the VPN name and the topology type directly from the name of the VRF.

Figure 1-4 shows a network in which two of the four sites are members of two VPNs, and illustrates which routes are included in the VRFs for each site.

Figure 1-4 VRFs for Sites in Multiple VPNs



VRF Implementation Considerations

When implementing VPNs and VRFs, Cisco recommends you keep the following considerations in mind:

- A local VRF interface on a PE is not considered a directly-connected interface in a traditional sense. When you configure, for example, a Fast Ethernet interface on a PE to participate in a particular VRF/VPN, the interface no longer shows up as a directly-connected interface when you issue a **show ip route** command. To see that interface in a routing table, you must issue a **show ip route vrf vrf_name** command.
- The global routing table and the per-VRF routing table are independent entities. Cisco IOS commands apply to IP routing in a global routing table context. For example, `show ip route`, and other EXEC-level show commands—as well as utilities such as **ping**, **tracert**, and **telnet**—all invoke the services of the Cisco IOS routines that deal with the global IP routing table.
- You can issue a standard Telnet command from a CE router to connect to a PE router. However, from that PE, you must issue the following command to connect from the PE to the CE:

```
telnet CERouterName /vrf vrf_name
```

 Similarly, you can utilize the **Tracert** and **Ping** commands in a VRF context.
- The MPLS VPN backbone relies on the appropriate Interior Gateway Protocol (IGP) that is configured for MPLS, for example, EIGRP or OSPF. When you issue a **show ip route** command on a PE, you see the IGP-derived routes connecting the PEs together. Contrast that with the **show ip route vrf VRF_name** command, which displays routes connecting customer sites in a particular VPN.

VRF Configuration Commands

The configuration commands to create a VRF instance are as follows:

Step	Command	Description
1	Router# configure terminal Router(config)#	Enter global configuration mode.
2	Router(config)# ip vrf vrf_name	For example, <code>ip vrf CustomerA</code> initiates a VPN routing table and an associated CEF table named CustomerA. The command enters VRF configuration submode to configure the variables associated with the VRF.
3	Router(config-vrf)# rd RD_value	Enter the eight-byte route descriptor (RD). The PE prepends the RD to the IPv4 routes prior to redistributing the route into the MPLS VPN backbone.
4	Router(config-vrf)# route-target import export both community	Enter the route-target information for the VRF.

For detailed information about these configuration commands, refer to Appendix B, “Cisco MPLS VPN Solutions Command Reference.”

Route Distinguishers and Route Targets

MPLS-based VPNs employ BGP to communicate between PEs to facilitate customer routes. This is made possible through extensions to BGP that carry addresses other than IPv4 addresses. A notable extension is called the *route distinguisher* (RD).

The purpose of the route distinguisher (RD) is to make the prefix value unique across the backbone. Prefixes should use the same RD if they are associated with the same set of route targets (RTs) and anything else that is used to select routing policy. The community of interest association is based on the route target (RT) extended community attributes distributed with the Network Layer Reachability Information (NLRI). The RD value must be a globally unique value to avoid conflict with other prefixes.

The MPLS label is part of a BGP routing update. The routing update also carries the addressing and reachability information. When the RD is unique across the MPLS VPN network, proper connectivity is established even if different customers use non-unique IP addresses.

For the RD, every CE that has the same overall role should use a VRF with the same name, same RD, and same RT values. The RDs and RTs are *only* for route exchange between the PEs running BGP. That is, for the PEs to do MPLS VPN work, they have to exchange routing information with more fields than usual for IPv4 routes; that extra information includes (but is not limited to) the RDs and RTs.

The route distinguisher values are chosen by the MPLS VPN Solution software.

- CEs with hub connectivity use `bgp_AS:value`.
- CEs with spoke connectivity use `bgp_AS:value + 1`

Each spoke uses its own RD value for proper hub and spoke connectivity between CEs; therefore, the MPLS VPN Solution software implements a new RD for each spoke that is provisioned.

MPLS VPN Solution chooses route target values by default, but you can override those values if necessary.

Route Target Communities

The mechanism by which MPLS VPN controls distribution of VPN routing information is through the VPN route-target extended MP-BGP communities. An extended MP-BGP community is an eight octet structure value. MPLS VPN uses route-target communities as follows:

- When a VPN route is injected into MP-BGP, the route is associated with a list of VPN route-target communities. Typically, this is set through an export list of community values associated with the VRF from which the route was learned.
- An import list of route-target communities is associated with each VRF. This list defines the values that should be matched against to decide whether a route is eligible to be imported into this VRF.

For example, if the import list for a particular VRF is {A, B, C}, then any VPN route that carries community value A, B, or C is imported into the VRF.

CE Routing Communities

A VPN can be organized into subsets called *CE routing communities*, or CERCs. A CERC describes how the CEs in a VPN communicate with each other. Thus, CERCs describe the logical topology of the VPN. MPLS VPN Solution can be employed to form a variety of VPN topologies between CEs by building hub and spoke or full mesh CE routing communities. CERCs are building blocks that allow you to form complex VPN topologies and CE connectivity.

The most common types of VPNs are *hub-and-spoke* and *full mesh*.

- A hub-and-spoke CERC is one in which one or a few CEs act as hubs, and all spoke CEs talk only to or through the hubs, never directly to each other.
- A full mesh CERC is one in which every CE connects to every other CE.

These two basic types of VPNs—full mesh and hub and spoke—can be represented with a single CERC.

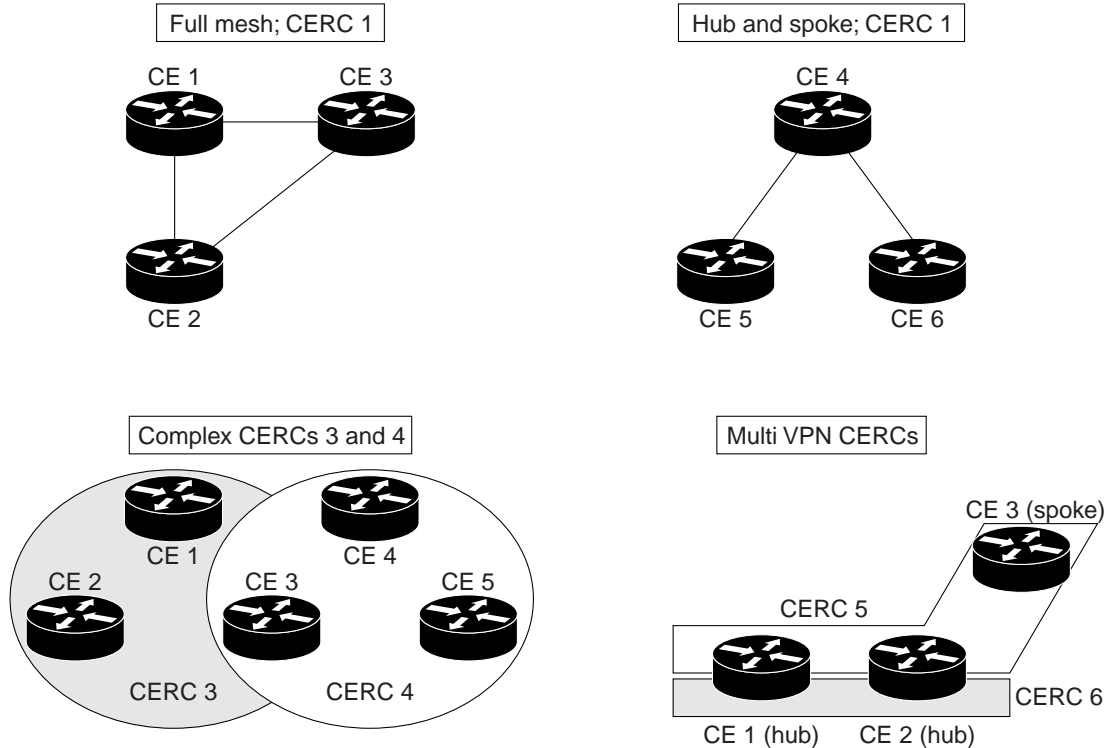
Whenever you create a VPN, the MPLS VPN Solution software creates one default CERC for you. This means that until you need advanced customer layout methods, you will not need to define new CERCs. Up to that point, you can think of a CERC as standing for the VPN itself—they are one and the same. If, for any reason, you need to override the software's choice of route target values, you can do this by editing the CERC definition, since this is where these values are stored.

To build very complex topologies, it is necessary to break down the required connectivity between CEs into groups, where each group is either fully meshed, or has a hub and spoke pattern. (Note that a CE can be in more than one group at a time, so long as each group has one of the two basic patterns.) Each subgroup in the VPN needs its own CERC. Any CE that is only in one group just joins the corresponding CERC (as a spoke if necessary). If a CE is in more than one group, then you can use the Advanced Setup choice during provisioning to add the CE to all the relevant groups in one service request. Given this information, the provisioning software does the rest, assigning route target values and VRF tables to arrange exactly the connectivity the customer requires. You can use the Topology tool to double-check the CERC memberships and resultant VPN connectedness.

MPLS VPN Solution supports multiple CEs per site and multiple sites connected to the same PE. Each CERC has unique route targets (RT), route distinguisher (RD) and VRF naming. After provisioning a CERC, it is a good idea to run the audit reports to verify the CERC deployment and view the topologies created by the service requests. The product supports linking two or more CE routing communities in the same VPN.

Figure 1-5 shows several examples of the topologies that MPLS VPN Solution CERCs can employ.

Figure 1-5 Examples of CERC Topologies



Hub and Spoke Considerations

In hub-and-spoke MPLS VPN environments, the spoke routers have to have unique Route Distinguishers (RDs). In order to use the hub site as a transit point for connectivity in such an environment, the spoke sites export their routes to the hub. Spokes can talk to hubs, but spokes never have routes to other spokes.

Due to the current MPLS VPN implementation, you must apply a different RD for each spoke VRF. The MP-BGP selection process applies to all the routes that have to be imported into the same VRF plus all routes that have the same RD of such a VRF. Once the selection process is done, only the best routes are imported. In this case this can result in a best route which is not imported. Thus, customers must have different RDs per spoke-VRF.

Full Mesh Considerations

Each CE Routing Community (CERC) has two distinct RTs: a hub RT and a spoke RT. When building a full mesh topology, always use the hub RT. Thus, when a need arises to add a spoke site for the current full mesh topology, you can easily add the spoke site without reconfiguring any of the hub sites. The existing spoke RT can be used for this purpose. This is a strategy to prevent having to do significant reprovisioning of a full mesh topology to a hub-and-spoke topology.

Quality of Service and Class of Service

Quality of Service (QoS) and Class of Service (CoS) enable the service provider to offer differentiated IP-based service levels and tiered pricing. QoS refers to the ability of a network to provide better service to selected network traffic. In particular, QoS features provide better and more predictable network service by the following:

- Supporting dedicated bandwidth
- Improving loss characteristics
- Avoiding and managing network congestion
- Setting traffic priorities across the network

CoS refers to the methods that provide *differentiated service*, in which the network delivers a particular kind of service based on the class of service specified for each packet. CoS provides specific categories of service such as Gold, Silver, and Best-Effort service classes.

To properly deploy QoS, enforcement of QoS measurements and policies must be in place throughout the network, from the first internetwork forwarding device (such as a Layer 2 switch or router) to the last device that front-ends the ultimate IP destination station. QoS requires an end-to-end approach because it requires mechanisms both at the edge and in the core.

To service providers, QoS is desirable because it has the potential of helping them support many types of traffic (data, voice, and video) over the same network infrastructure. It allows them to offer business-quality IP VPN services, and the end-to-end service level agreements (SLAs) that customers demand.

In an MPLS VPN environment, the service provider must consider both packet and cell routers. In a packet environment, MPLS Class of Service is fairly straightforward. A PE simply copies the IP precedence to the MPLS Class of Service field. The CoS field can then be used as input to Weighted Random Early Detection (WRED), as well as Weighted Fair Queuing (WFQ). The challenge is to provide MPLS CoS in environments where PEs are connected to ATM. Class of Service is more involved on ATM interfaces and within the ATM PEs themselves. QoS is discussed in-depth in other resources available from Cisco. The emphasis in this section is to investigate differentiated services in MPLS intranet and extranet VPN environments.

In mega-scale VPNs, applying QoS on a flow-by-flow basis is not practical because of the number of IP traffic flows in carrier-sized networks. The key to QoS in large-scale VPNs is implementing controls on a set of service classes that applications are grouped into. For example, a service provider network may implement three service classes: a high-priority, low-latency “premium” class; a guaranteed-delivery “mission-critical” class; and a low-priority “best-effort” class. Each class of service is priced appropriately, and subscribers can buy the mix of services that suits their needs. For example, subscribers may wish to buy guaranteed-delivery, low-latency service for their voice and video conferencing applications, and best-effort service for e-mail traffic and bulk file transfers.

Because QoS requires intensive processing, the Cisco model distributes QoS duties between the PEs and core routers. This approach assumes a lower-speed, high-touch edge and a high-speed, lower-touch core for efficiency and scalability.

Cisco IOS QoS/CoS Toolkit

Cisco IOS software includes several Layer 3 QoS features that are particularly applicable to VPN provisioning and management. MPLS-enabled networks make use of the following Cisco IOS QoS features to build an end-to-end QoS architecture:

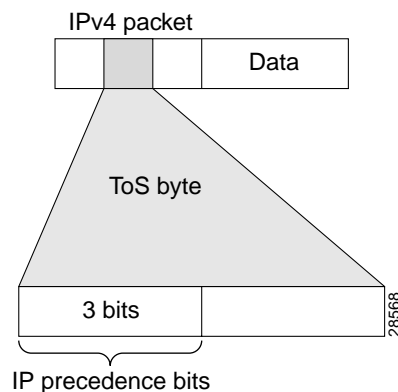
- IP Precedence

- Committed Access Rate (CAR)
- Weighted Random Early Detection (WRED)
- Weighted Fair Queuing (WFQ)
- Generic Traffic Shaping

IP Precedence

IP Precedence utilizes the three precedence bits in the IPv4 header Type-of-Service field to specify class of service for each packet, as shown in Figure 1-6. You can partition traffic in up to six classes of service using IP Precedence (two others are reserved for internal network use). Queuing technologies throughout the network can use this signal to provide the appropriate expedited handling.

Figure 1-6 Type of Service Field in the IP Precedence Header



Committed Access Rate (CAR)

Committed Access Rate (CAR) is Cisco's traffic policing tool for instituting a QoS policy at the edge of a network. CAR allows you to identify packets of interest for classification with or without rate limiting. CAR allows you to define a traffic contract in routed networks. You can classify and police traffic on an incoming interface, and set policies for handling traffic that exceeds a certain bandwidth allocation. CAR can be used to set IP precedence based on extended access list classification. This allows considerable flexibility for precedence assignment, including allocation by application, port, source address, destination address, and so on. As a rule-based engine, CAR classifies traffic based on flexible rules, including IP precedence, IP access lists, incoming interface, or MAC address. It limits the rate to the defined ingress thresholds to help allay congestion through the core.

For monitoring details, see the "Monitoring Data with the Committed Access Rate MIB" section on page 5-18.

MPLS VPN Solutions software uses Committed Access Rate on the ingress or egress PE interfaces to perform two main tasks:

- Classify traffic into a maximum of four classes of service
- Rate limit traffic to a specified bandwidth

The PE can rate limit traffic to the subscribed bandwidth and mark the traffic that is within the specified bandwidth as *in-contract*, and mark traffic above the specified bandwidth as *out-of-contract*.

Marking a packet as in-contract or out-of-contract is done by setting the first bit of the precedence bits in the IP header. The appropriate class is indicated by the remaining two precedence bits (see). Traffic that exceeds any class is marked as out-of-contract, and this traffic can be dropped or mapped to a lower class of service. The out-of-contract bandwidth is initially set to the in-contract bandwidth, but you can set this in the product's VPN Console to the values appropriate for the customer (see the "Defining a Class of Service Profile" section on page 3-23).

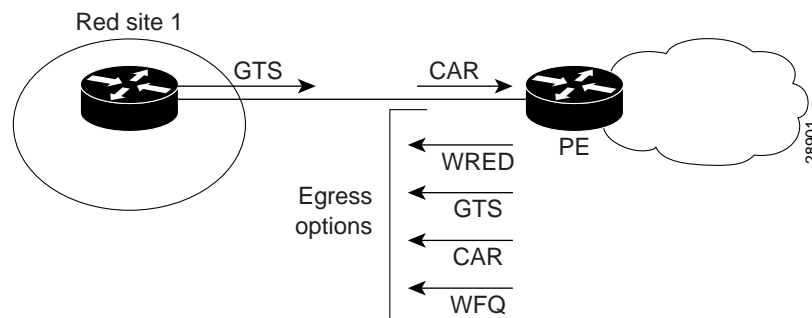
Table 1-1 Mapping IP Precedence to Class of Service

IP Precedence	Contract Status	Class of Service
111	In-contract	Class 1
110	In-contract	Class 2
101	In-contract	Class 3
100	In-contract	Class 4
011	Out-of-contract	Class 1
010	Out-of-contract	Class 2
001	Out-of-contract	Class 3
000	Out-of-contract	Class 4

The customer can initially "paint" the packets that leave the customer edge router (heading to the PE), and MPLS VPN Solution allows policing or repainting of packets that enter the provider edge router.

Figure 1-7 shows GTS applied to the CE's egress interface (Red site 1) and CAR applied to the PE's ingress interface, as well as the CoS options available on the PE's egress interface.

Figure 1-7 CoS Options on the PE and CE



Generic Traffic Shaping (GTS)

Generic Traffic Shaping (GTS) shapes traffic by reducing outbound traffic flow to avoid congestion. It does this by constraining traffic to a particular bit rate using the token bucket mechanism. GTS applies on a per-interface basis and can use access lists to select the traffic to shape. GTS can also be applied to a specific access list on an interface.

GTS smooths out bandwidth peaks by class of service. GTS shapes by looking at bandwidth (for example, 40 kbps for "gold," or best of service), and uses a buffering technique for handling excess (or out-of-contract) traffic. GTS is useful as a tool to protect the link between the PE and CE by limiting bursty traffic that exits the CE.

Generic Traffic Shaping can take place on the CE's egress subinterface for traffic shaping and on the PE's egress subinterface for congestion management.

MPLS VPN Solution has the option to apply traffic shaping to either a) in-contract and out-of-contract packets leaving the CE interface, or b) out-of-contract traffic flow exiting the CE interface. GTS is provisioned so that traffic does not exceed a specified bandwidth and performs burst traffic control for the PE-CE link.

The out-of-contract bandwidth is, by default, equal to the in-contract bandwidth. You have the option to allow the out-of-contract bandwidth to be set to a percentage of the in-contract bandwidth. MPLS VPN Solution also allows the application of adaptive traffic shaping for Frame Relay; the traffic shaping responds to Frame Relay traffic congestion control.

Weighted Random Early Detection (WRED)

WRED provides congestion avoidance. This technique monitors network traffic load in an effort to anticipate and avoid congestion at common network bottlenecks, as opposed to congestion management techniques that operate to control congestion once it occurs.

WRED is designed to avoid congestion in internetworks before it becomes a problem. It leverages the flow monitoring capabilities of TCP. It monitors traffic load at points in the network and discards packets if the congestion begins to increase. The result is that the source detects the dropped traffic and slows its transmission. WRED interacts with other QoS mechanisms to identify class of service in packet flows. It selectively drops packets from low-priority flows first, ensuring that high-priority traffic gets through. WRED is supported on the same interface as WFQ. You should run both of these queueing algorithms on every interface where congestion is likely to occur. In the service provider core, apply WRED by IP precedence and WFQ by service class.

Weighted Fair Queuing (WFQ)

WFQ addresses situations where it is desirable to provide consistent response time to heavy and light network users alike without adding excessive bandwidth. WFQ is a flow-based queueing algorithm that does two things simultaneously: it schedules interactive traffic to the front of the queue to reduce response time, and it fairly shares the remaining bandwidth among lower-priority flows.

Weighted Fair Queuing (WFQ) employs a differential scheduling policy that results in packets of different classes getting different amounts of link bandwidth during outbound congestion. WFQ is the differentially-oriented counterpart to a first-in, first-out (FIFO) scheduling policy.

WFQ ensures that queues are not starved for bandwidth and that traffic achieves predictable service so that mission-critical traffic receives highest priority to ensure guaranteed delivery and latency. Lower-priority traffic streams share the remaining capacity proportionally among them. The WFQ algorithm also addresses the problem of round-trip delay variability. If multiple high-volume conversations are active, their transfer rates and inter-arrival periods are made much more predictable. Algorithms such as the Transmission Control Protocol (TCP) congestion control and slow-start features are much enhanced by WFQ. The result of WFQ is more predictable throughput and response time for each active flow.

WFQ is IP precedence-aware; that is, it is able to detect higher priority packets marked with precedence by the IP forwarder and schedule them faster, providing superior response time for this traffic. The IP precedence field has values between 0 (the default) and 7. As the precedence value increases, the algorithm allocates more bandwidth to that conversation to make sure that it gets served more quickly when congestion occurs. WFQ assigns a weight to each flow, which determines the transmit order for queued packets. It provides the ability to reorder packets and control latency at the edge and in the core.

By assigning different weights to different service classes, a switch can manage buffering and bandwidth for each service class. This mechanism constrains delay bounds for time-sensitive traffic such as voice or video.

Service providers can tailor each class to the specific service needs of their customers. For example, a service provider can offer a “Gold class” for voice traffic. Here, a large bandwidth allocation policy ensures that sufficient bandwidth is available for all the cells in the voice queue while a moderately-sized buffer limits the potential cell delay. Since these shares are relative weights, allocating a large share to gold means that a minimum is guaranteed. If the gold class is under utilized, the bandwidth is shared by the remaining classes in proportion to their weights. This ensures maximum efficiency and that paying customer traffic is sent if bandwidth is available.

Proper QoS/CoS Placement in the Network

QoS/CoS application is easy to implement in a non-ATM MPLS environment. When QoS must be implemented in an end-to-end fashion, two areas of implementation need to be looked at— the ingress and egress edges of the network and the core network.

At the edges of the network, traffic enforcement and policing need to be present. Therefore, at the edges of the network, Cisco’s Committed Access Rate (CAR) is required.

In the core of the network, techniques such as WRED and WFQ need to be considered.

The IP precedence setting, if policy calls for it, is modified at the ingress of a network. It is also possible, for certain environments, to adjust the IP precedence field at the egress of the network.

NetFlow Collector and MPLS VPN Solution Software

Although NetFlow is not part of the MPLS VPN Solution software suite, it is an important element in the MPLS VPN network scheme, and operators are required to specify the NetFlow Collector devices in the service provider’s network (see the “Specifying the NetFlow Collector Devices in the Network” section on page 5-3). Cisco recommends that service providers schedule MPLS VPN Solution to collect data from NetFlow Collector devices every three hours.

The NetFlow data is stored on the NetFlow workstations in binary flat files. Because NetFlow sends data from the router in User Datagram Protocol (UDP) packets, Cisco recommends that the NetFlow Collector 3.0 device be located on a LAN connected directly to the PE or the Management PE (MPE) device.

NetFlow services consist of high-performance IP switching features that capture a rich set of traffic statistics exported from routers and switches while they perform their functions. The exported NetFlow data consists of traffic *flows*, which are unidirectional sequences of packets between a particular source device and destination device that share the same protocol and transport-layer information. The captured traffic statistics can be used for a wide variety of purposes, such as network analysis and planning, network management, accounting, billing, and data mining.

Because of their unidirectional nature, flows from a client to a server are differentiated from flows from the server to a client. Flows are also differentiated on the basis of protocol. For example, Hypertext Transfer Protocol (HTTP) Web packets from a source device to a destination device constitute a separate flow from File Transfer Protocol (FTP) packets between the same pair of devices.

NetFlow works by recording the initial IP packet attributes in a flow such as IP protocol type, type of service (ToS), and interface identifiers. In this way, NetFlow enables subsequent packets that belong to the same flow to be efficiently matched and counted. Likewise, it streamlines services appropriate to

that traffic flow, such as security filtering, quality of service (QoS) policy, or traffic engineering. This real-time information flow is held in the NetFlow Cache, where it can be retrieved with a NetFlow data export operation.

A Cisco IOS service called NetFlow Policy Routing (NPR) integrates policy routing with NetFlow services. Because it is supported in conjunction with the Cisco Express Forwarding (CEF) architecture, NPR's performance can be scaled across the distributed platforms.

NetFlow FlowCollector 3.0 provides fast, scalable, and economical data collection from multiple NetFlow Export-enabled devices. A UNIX or Windows NT application, the FlowCollector reduces data volume through selective filtering and aggregation. The FlowCollector also stores flow information in flat files on disk for post-processing.

The NetFlow Data Analyzer is a network traffic analysis tool that works with the NetFlow FlowCollector. The Data Analyzer produces graphs, sorts, and builds tables from the data sets exported by NetFlow FlowCollector 3.0.