Layer 3 MPLS VPN Enterprise Consumer Guide Version 2

This document is written for networking engineers and administrators responsible for implementing a Layer 3 (L3) MPLS VPN service from a service provider (SP) network. It describes important considerations when choosing an SP and making the necessary connections. This document outlines these considerations, but it is not meant to be a comprehensive design guide.

Note Throughout this document, references to MPLS VPN mean Layer 3 MPLS VPN. The terms "self-managed" and "unmanaged" service are synonymous and refer to a service in which the enterprise customer is responsible for managing the CE devices as well as the devices within each of their sites. Also, the terms "customer" and "enterprise" are also synonymous and refer to the subscriber of the MPLS VPN service.

Contents

Corporate Headquarters:
Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134-1706 USA

MPLS VPN Primer

VPN service offers a cost effective way to expand geographically or to replace expensive dedicated circuits such as leased lines, Frame Relay, or ATM networks. An L3 MPLS VPN service is an attractive option because it provides full mesh capabilities and more bandwidth in the WAN in a more cost-effective manner than legacy TDM or packet switching technologies such as Frame Relay.

This section provides an MPLS VPN primer for enterprise customers looking for a quick introduction to this service. This section includes the following topics:

- Layer 3 MPLS VPN Services Introduction, page 3
- Layer 3 MPLS VPN Operation, page 7
- Strengths and Limitations of MPLS Layer 3 VPN Services, page 6
- Layer 3 MPLS VPN Operation, page 7

Layer 3 MPLS VPN Services Introduction

L3 MPLS VPN services allow businesses to outsource their current network core using a private IP-based service offering from a service provider. Unlike current overlay networks (such as ATM or Frame Relay service offerings), MPLS VPNs require that the enterprise peer with the SP at the IP L3 level. In this scenario, the SP network is involved in the L3 routing of the IP packets delivered by the enterprise.

This capability is implemented through Virtual Routing/Forwarding (VRF) tables for each customer, and MPLS labels to de-multiplex and to tunnel customer traffic through the SP core. Because the SP network participates in the routing of customer traffic, each enterprise must inject its prefixes into the appropriate VRF table in the SP network. The SP is responsible for ensuring that these routes are distributed to the appropriate customer VRF tables.

Routing scenarios can sometimes be complex, such as in a customer hub-and-spoke topology where traffic to and from each spoke is routed through the hub. However, the most common deployment is an any-to-any topology where any customer device can connect directly to the L3 MPLS VPN. Enterprise traffic entering the SP domain is then routed based on the information in the VRF table and encapsulated with MPLS labels to ensure proper tunneling and de-multiplexing through the core.

Layer 3 MPLS VPN Terminology

Figure 1 illustrates many of the acronyms and terms used when discussing L3 MPLS VPNs.

Figure 1 MPLS Layer 3 VPN Component Terminology

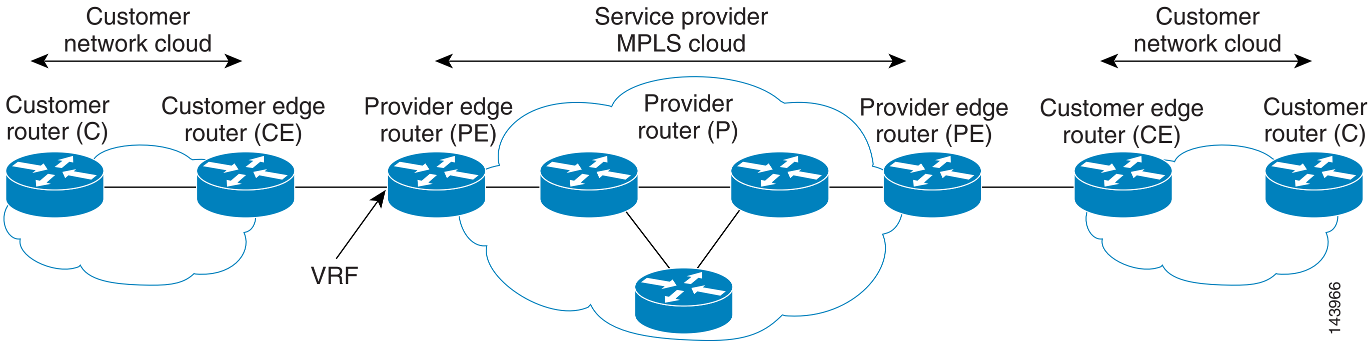

Customer network cloud
Service provider MPLS cloud
Customer network cloud
Customer router (C)
Customer edge router (CE)
Provider edge router (PE)
Provider router (P)
Provider edge router (PE)
Customer edge router (CE)
Customer router (C)
VRF
143966

Table 1 defines the acronyms and terms you should understand when designing and implementing an L3 MPLS VPN.

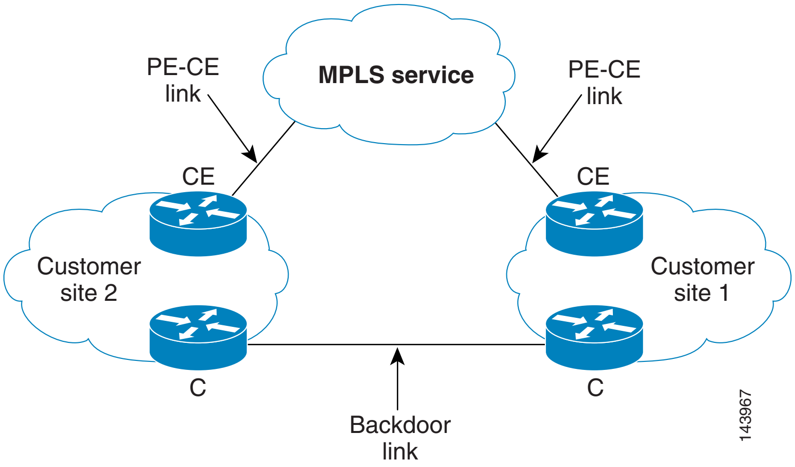
Table 1 L3 MPLS VPN Terminology

Term	Meaning
Backdoor connectivity	Either a dynamic or permanent link, outside of the MPLS VPN cloud, over which a routing adjacency is formed to pass routing information that ties two customer domains together. This link is typically used to connect two geographically distinct sites and usually runs the same IGP protocol as the customer site. An example of a backdoor link is illustrated in Figure 2.
C	Customer router that is connected only to other customer devices.
CE	Customer edge router that peers at Layer 3 to the provider edge. The PE-CE interface runs either a dynamic routing protocol (eBGP, RIPv2, EIGRP, or OSPF) or a static routing protocol (Static, Connected).
Global routing/ forwarding table	The non-VRF routing and forwarding table used in the SP core for infrastructure addressing reachability.
Label	In this document, this refers to an MPLS frame-based label.

Table 1 L3 MPLS VPN Terminology (continued)

Term	Meaning
MP-BGP	Multi-Protocol Border Gateway Protocol. In an MPLS VPN context, this protocol is run between PE routers to exchange customer prefixes in a VPNv4 format.
Managed CE service	Some service providers may offer an added service along with the Layer 3 MPLS VPN offering known as a managed CE service. The SP handles the operations, management, and administration of the CE router at one or more sites. There are typically added charges for what is essentially outsourced management of the CE devices.
P	Provider router, which resides in the core of the provider network. In an MPLS VPN context, the P router participates in the control plane for customer prefixes. The P router is sometimes referred to as a label switch router (LSR), in reference to its primary role in the core of the network, performing label switching/swapping of MPLS traffic.
PE	Provider edge router. The PE router sits at the edge of the MPLS SP cloud. In an MPLS VPN context, separate VRF routing tables are allocated for each user group. Also, the PE still contains a global routing table for routes in the core SP infrastructure. The PE is sometimes referred to as a label edge router (LER) or edge label switch router (ELSR) in reference to its role at the edge of the MPLS cloud, performing label imposition and disposition.
RD	Route distinguisher, which is a 64-bit value defined uniquely for each user group. The RD is combined with the customer IPv4 prefix to guarantee that the resulting VPNv4 prefix is unique.
RT	Route target, which is a 64-bit value used as a BGP extended community attribute. The RT is used to determine the VPNv4 routes that should be installed in the respective VRF tables.
VPNv4	The combination of the RD and customer IPv4 prefix. These VPNv4 prefixes are passed in MP-BGP.
VRF	The virtual routing and forwarding table, which is separate from the global routing table that exists on PE routers. Routes are injected into the VRF from the CE-PE routing protocols for that VRF and any MP-BGP announcements that match the defined VRF route targets (RTs).

Figure 2 Backdoor Link Example



Strengths and Limitations of MPLS Layer 3 VPN Services

MPLS Layer 3 VPN services offer several advantages, including flexibility, scalability, and cost reduction. Table 2 lists some of the high-level advantages and disadvantages of the service. For more detailed information, see the following URL: http://www.cisco.com/en/US/products/ps6557/products_ios_technology_home.html

Table 2 Advantages and Disadvantages of MPLS Layer 3 VPN Services

Advantages	Disadvantages
Scalable routing model—The Layer 3 peer-to-peer model reduces the demands on the CE device (low CPU trend, less IDB, and so forth). This is an improvement over the overlay model of a traditional Layer 2 SP offering (ATM and Frame Relay).	IP only—L3 MPLS VPNs transport only IPv4 traffic. Non-IP protocols need to be tunneled through some mechanism (such as GRE) on the CE or C device before reaching the PE.
Scalable bandwidth model—A Layer 3 MPLS VPN model is not limited by the PE-CE media type, but is limited only by the SP infrastructure for PE-CE (for example, Frame Relay, POS, or GE).	SP dependency—The customer is dependent on the SP in regards to Layer 3 features and capabilities. For example, although Cisco offers IP Multicast as a feature for MPLS VPNs (mVPN), not every SP offers it as a service. Layer 3-based convergence and QoS capabilities are also dependent on the SP offering, and SLAs must be negotiated to manage these requirements.

Table 2 Advantages and Disadvantages of MPLS Layer 3 VPN Services (continued)

Advantages	Disadvantages
Reduced total cost of ownership—MPLS cost is lower compared to other solutions because of outsourced networking responsibility and lower service costs (typically 10–40 percent lower).	Possible difficulties in integration—The difficulty of integration from Layer 2 to Layer 3 peering varies greatly depending on the SP offering. For example, EIGRP as a PE-CE protocol is ideal for customers already running EIGRP as their IGP. However, if the SP does not offer this service, integration with a different routing protocol, such as eBGP, might require design changes and training of network staff.
Intelligent QoS—The SP can now provide L3 QoS, which allows more intelligence in the SP core compared to L2 QoS.	
Any-to-any connectivity—By peering with the SP at Layer 3, each site (after it is terminated into the SP cloud) can be configured with IP route reachability to all other customer sites. This allows any-to-any connectivity and offers more efficient routing compared to ensuring connectivity between spokes in a traditional hub-and-spoke topology. This is an important advantage where there is a growing trend toward distributed applications and VoIP.	

Layer 3 MPLS VPN Operation

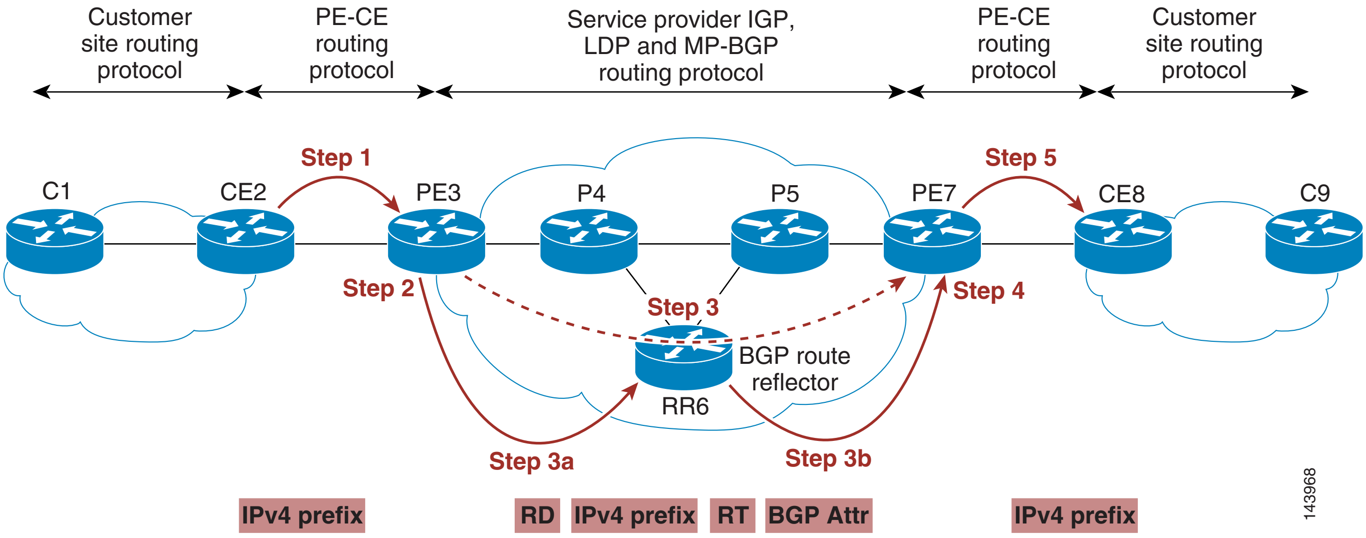
This section briefly examines the L3 MPLS VPN control and data planes, and includes the following topics:

- Layer 3 MPLS VPN Route Distribution Operation, page 7
- Layer 3 MPLS VPN Forwarding Operations, page 8

Layer 3 MPLS VPN Route Distribution Operation

Figure 3 illustrates an example of BGP VPN route distribution using MP-BGP between a VPN that terminates on PE3 and PE7. The customer devices (C1 and CE2 on the left, and CE8 and C9 on the right) participate in the same VPN.

Figure 3 Figure 3BGP VPN Route Distribution



The distribution steps are as follows:

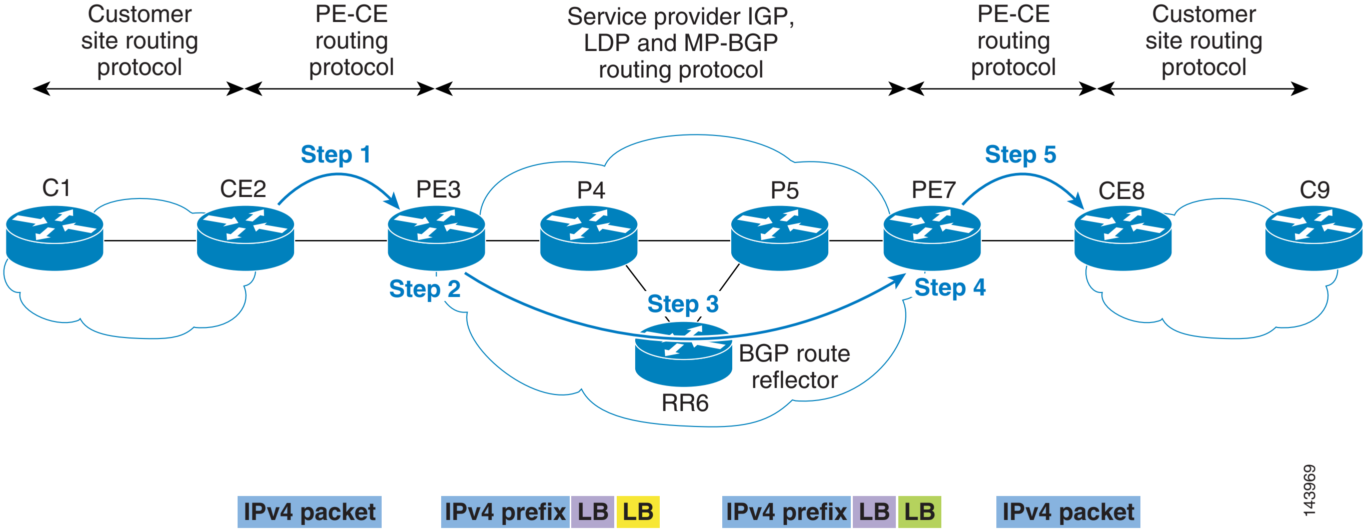
1. Customer routes are injected into the VRF table at PE3 using static, RIPv2, OSPF, or BGP routing protocol between the PE and the CE. The customer routes are passed as IPv4 prefixes (shown in the red shaded box under Step 1).
2. At PE3, the routes in the customer VRF are exported into MP-BGP as VPNv4 prefixes. To ensure VPNv4 route uniqueness, the customer IPv4 routes are prepended with a uniquely defined RD to create a distinct VPNv4 prefix. Every VRF configuration requires an RD to be defined. Its uniqueness guarantees customer VPNv4 uniqueness.
3. The exported routes are sent across the MPLS backbone between the BGP peers in PE3 and PE7. This process repeats for any other BGP peers that have members in the same VPN. Note that this step shows a logical connection between the two BGP peers. There can be a series of BGP route reflectors in between performing the VPN distribution as shown in Steps 3a and 3b.

   The VPNv4 prefix (shown in red shaded boxes under Step 3) is composed of the RD and the customer IPv4 prefix. Because this VPNv4 prefix is a BGP route, multiple mandatory and optional BGP attributes are carried along with the prefix. One of these attributes is the route target (RT), which is an extended community BGP attribute.
4. The routes are imported into the correct VRF at PE7. Every VRF configuration contains VRF import and export definitions. The export definitions define which RTs are attached to the BGP VPNv4 prefix, as described in Step 3. The export definitions define the RTs that are carried along with the VPNv4 prefix on export. The import definitions define the RT tagged prefixes that are imported into the VRF. Only VPNv4 prefixes with a matching RT tag to the VRF import RT definitions are imported into that VRF.
5. The routes are accessible from a VPN at each site.

Layer 3 MPLS VPN Forwarding Operations

Figure 4 illustrates the process of packet forwarding for a packet originating from the customer cloud containing C1 and CE2 to the far-end customer cloud containing CE8 and C9.

Figure 4 *MPLS Data Forwarding Example*



1. The customer cloud composed of C1 and CE2 originates an IPv4 packet destined to an address at the far end (CE8 and C9). The routing entry on CE2 for the destination prefix forwards the packet to the PE3 device.
2. PE3 receives the customer packet and does a routing lookup according to the VRF table that is bound to that interface. In this case, the route resolves to a BGP prefix originated from PE7. PE3 imposes two labels on the IPv4 packet. The first label, referred to in this document as the VPN label, (shown in the purple "LB" shaded box) is the label that is used to uniquely identify a customer VPN prefix. The second label, referred to in this document as the forwarding label (shown by the yellow "LB" shaded box) is the label used to tunnel the packet through the P core to the far-end PE7 device.
3. The labeled packet is now forwarded at each hop through the SP core. Each P router makes a forwarding decision based on the top level label, and this top level label is swapped with a new label. This is shown by the yellow "LB" shaded box, and the outgoing packet is shown with a green "LB" shaded box. The underlying packet and inner label are left undisturbed during this process.
4. Eventually, PE7 receives the labeled packet and recognizes the inner VPN label (purple "LB") as a VPN label for that specific customer prefix. The VPN label is stripped and a forwarding decision for the IPv4 packet is made based on the VPN label.

   P5 may remove the top level label, leaving only the inner label when forwarding to PE7. This concept is known as penultimate hop popping (PHP), where the penultimate hop removes the top level label. The relevance to the enterprise is that in a PHP scenario, the SP-marked EXP value may not be copied down to the inner label. This depends on the MPLS QoS mode chosen. This is relevant only if the traffic from the PE to the CE (for example, PE7 to CE8 in Figure 4) must be queued based on the SP EXP marking
5. The original IPv4 packet is forwarded by the switch to the appropriate customer VRF interface.

The MPLS label is a 32-bit shim that is inserted between the L2 data link header and the underlying payload (in this case an IPv4 packet). Figure 5 illustrates the format of the 32-bit label.

Figure 5 MPLS Label Detail

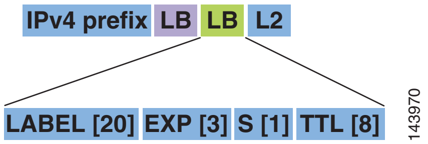


Table 3 describes each field in this label:

Table 3 MPLS Label Field Descriptions

Field ID	Length	Purpose
LABEL	20 bits	Allocated for the actual label value.
EXP	3 bits	MPLS experimental bits. A Cisco convention is to use these experimental bits as a means of representing the class of service (CoS) of the MPLS frame.
S	1 bit	End-of-stack (EOS) bit. Some MPLS applications such as L3 MPLS VPNs require the use of multiple labels. The EOS is set on the last label in the stack of labels.
TTL	8 bits	Time to live for the MPLS frame. This performs a similar function to an IPv4 TTL.

MPLS VPNs, unlike other VPN types such as IPsec, perform no encryption. Despite this, however, a Layer 3 MPLS VPN service offers equivalent security to that of an ATM/Frame Relay service offering through the use of distinct routing tables and label spoofing mechanisms.

Third-party verification of the security of MPLS can be found at a Miercom study at the following URL: http://www.mier.com/reports/cisco/MPLS-VPNs.pdf

For further information regarding MPLS security, see the following URL: http://www.cisco.com/en/US/tech/tk436/tk428/technologies_white_paper09186a00800a85c5.shtml

Choosing a Service Provider

When choosing an SP for MPLS VPN services, you must consider your business needs. There is no single best choice for every organization. The best choice is the provider or providers that best meet your organizational needs and that offer the most transparent service. For enterprise customers who have a Cisco Advanced Services (AS) contract a more exhaustive questionnaire is available through the local Cisco AS Network Consulting Engineer (NCE). Enterprise customers without an AS contract should contact their Services Account Manager (SAM).

Note A critical prerequisite before choosing an SP is assessing your business requirements, environment, and objectives. Invest the time to understand your network, its underlying infrastructure, and application needs. You should also know the network and application requirements of branch networks and other remote locations.

This section describes some criteria to consider when selecting a provider, and includes the following topics:

- General Architecture and Services, page 11
- Routing Considerations, page 15
- Layer 2 Access to the MPLS VPN Service, page 16
- QoS Capabilities, page 17
- Multicast Capabilities, page 17
- Security, page 18

General Architecture and Services

This section describes the general architecture and services you should consider when selecting an SP. It includes the following topics:

- Cisco Powered Networks, page 11
- Coverage, page 12
- Inter-AS MPLS VPN, page 12
- PE-CE IP Addressing, page 12
- Hub-and-Spoke Topology Considerations, page 12
- Extranet Support, page 13
- Remote Access and IPsec, page 13
- Backup Considerations, page 13
- Non-IP Application Support, page 13
- Managed CE Services, page 14
- SLA Agreement and Reporting, page 14

Cisco Powered Networks

A great starting point is to consider providers that are designated as Cisco Powered Networks. Service providers that display the Cisco Powered logo are uniquely positioned to help customers migrate to MPLS-based VPN services. These providers have earned the Cisco Powered designation by maintaining high levels of network quality and by basing their VPN services end-to-end on Cisco equipment.

In addition, an increasing number of Cisco Powered providers have earned the QoS Certification for VPN services. This means that they have been assessed by a third party for the ability of their SLAs to support real-time voice and video traffic, and for their use of Cisco best practices for QoS. Look for the QoS Certification as an extra indication that you can have confidence in the Cisco Powered provider.

Nearly 400 of the most successful service providers throughout the world offer services that have earned the Cisco Powered designation. Situated in 62 countries, these providers offer a wide range of services for both small and large businesses. From the basics, such as Internet access and web hosting, to managed services such as IP telephony and multiservice VPNs, these providers should be your first choice when you need to outsource a critical business function. For a list of recommended service providers, see the following URL: http://www.cisco.com/cpn

Coverage

Many companies need to expand their data networking to remote sites, data centers, or branch offices. Connectivity requirements may also span many regions in various countries. However, the services that specific providers offer may be limited geographically. Providers tend to offer more services in their home regions, and services are harder to obtain for remote regions. When evaluating L3 MPLS VPN services, you should understand the PE coverage and consider which cities around the world the PE routers used for customer connections are located. In many cases, providers have partners that provide local access. It is important to consider the locations where these partners provide PE routers, and to make sure this meets your organizational needs.

Inter-AS MPLS VPN

To establish a larger global footprint, providers may establish partnerships with other service providers to interconnect MPLS VPN networks. This is known as an interprovider MPLS VPN.

However, inter-AS MPLS VPNs can affect the availability or behavior of services such as QoS and Multicast. One provider may support these services in a different manner than the other, or a provider might not support a service at all. Therefore, it is important to consider SP inter-AS agreements and whether the implementation supports your network requirements.

PE-CE IP Addressing

Whether the MPLS VPN service is a managed CE service or not, the customer and the provider must agree about IP addressing on the CE-PE links. The service provider typically assumes the responsibility for determining the addresses to use. The SP may approach the address assignment in various ways, including the following:

- Private address space (RFC 1918)—In this scenario, addresses must be carefully assigned to prevent conflicts with RFC 1918 addresses used by the customer.
- Unnumbered addressing on the link—Although this may seem to be a good approach to save on address space, this approach causes a problem for network management devices, which are not able to capture a view of the PE-CE link. The use of unnumbered addressing requires the use of other addresses assigned to interfaces in the same routing table. This requires additional loopback interfaces for each VRF on the PE routers.
- SP address space—This allows each PE-CE link to have unique addresses but may require a large amount of address space.
- Customer address space—This also allows for each PE-CE link to be addressed uniquely. However, this may get complex if the address space used by the customer is RFC 1918 address space that overlaps with RFC 1918 addresses used by the SP. The SP may be required to configure their management devices to deal with overlapping addresses.

Whatever approach is taken for assigning PE-CE addresses, careful coordination between the SP and the customer is essential. Otherwise, IP connectivity issues or network management problems may occur.

Hub-and-Spoke Topology Considerations

Layer 2 WANs were often designed in a hub-and-spoke topology because of historical costs and capability constraints. In this topology, spoke sites are not able to communicate with each other directly and can communicate with each other only through the hub site.

Customers may wish to maintain a hub-and-spoke model even after converting to an MPLS VPN. SP implementations of hub-and-spoke MPLS VPNs can force spoke site traffic to route through a centralized hub site to reach other spoke sites. Such routing behavior may be critical for centralized services such as firewalling spoke-to-spoke traffic. However, because MPLS VPNs typically offer any-to-any connectivity, creating a hub-and-spoke topology adds a level of complexity to the service.

Extranet Support

Extranet support involves importing routes from one VRF into a different VRF that may service a different VPN site. Extranet VPNs support dynamic connectivity between your network and other networks subscribed to the same provider. This could be helpful for creating extranets with partners or vendors.

Remote Access and IPsec

Remote access to the MPLS VPN lets service providers extend services to the last mile using a broad range of access options, including dial-up, DSL, and cable technologies. This lets remote users securely access the corporate intranet and extranet using an MPLS VPN.

Customers with remote workers should consider whether the SP offers remote access to the MPLS VPN. The customer may also be interested whether the solution allows IPsec termination for connecting to the customer network. SPs that offer dial access or IPsec termination into the customer network can be used for outsourcing support for existing dial-up and remote office telecommuters.

Backup Considerations

You should also consider how the SP protects against primary MPLS VPN connectivity failures. Some L3 MPLS offerings may include a backup service that terminates in the customer VRF. Other offerings may provide an external leased line, in which case it is not integrated into the VRF.

In the latter case, or in cases where no backup is provided, the customer must implement their own backup mechanism (leased line, DMVPN, second provider), and a backdoor connection may be required. When using a backdoor connection, it is critical to understand how your backup mechanism works to avoid potential routing loops or sub-optimal routing.

Non-IP Application Support

When choosing to move to an MPLS VPN environment, customers must consider any legacy applications, such as SNA or DECnet, that they are required to support. Because the MPLS VPN architecture supports only IP traffic, how the SP provides non-IP traffic support is critical when legacy applications must be supported.

The SP may require the customer to maintain the existing Frame Relay or ATM network for legacy applications. On the other hand, the provider can support legacy applications using GRE tunneling to facilitate transport over the MPLS VPN network. GRE tunneling adds a layer of complexity to the architecture that may be best handled by having the SP manage the CE routers. This places responsibility for configuration and maintenance with the SP.

Managed CE Services

Businesses that move to MPLS VPNs can often choose to purchase a “managed” CE service from an SP, which can handle part or all of the requirements for installation, provisioning, management, and security of network equipment. Managed services provide enterprise customers immediate access to the benefits of an MPLS network, with network availability and security being managed by the SP.

With a managed CE service, it is important to understand the limits of administrative control of the CE device. Customers may not be allowed to make any changes to the CE router, or there may be feature restrictions placed on the managed CE. If so, you should know the turnaround time for necessary changes or features to be deployed by the SP. You should also understand the visibility provided into the router because this affects your ability to troubleshoot network problems.

SLA Agreement and Reporting

Everything described in Choosing a Service Provider, page 10, can potentially be included or negotiated in a service level agreement (SLA). The purpose of this subsection is to discuss SLAs in general.

An SLA sets the expectations between the provider and the customer. As an MPLS VPN customer, look for an SLA that answers the questions that are important to you, which may include the following:

- What is the provider is committed to deliver?
- How will the provider deliver on commitments?
- What is meant by network availability? Is it CE to CE, PE to PE or CE to PE?
- How are network performance agreements defined and measured? For example, is latency measured from CE to CE or PE to PE?
- Are any monitoring tools offered by the SP?
- What happens if a provider fails to deliver as promised?
- How quickly does the SP respond to network problems?
- How quickly do they respond to business growth needs by adding sites or equipment?

SLAs should not be limited to network performance and availability, but should encompass support and growth.

The details of an SLA may vary, but it should meet the specific requirements and network application needs of the customer. The following are some examples of network performance deliverables that might be negotiated:

- Bandwidth
- Latencies
- Jitter
- Packet drop
- Network availability
- SLA reporting

SLAs should be based on realistic and measurable commitments. Having the ability to measure against the commitments ensures the success of the agreement. Defining what should be measured, how and when it should be measured, and how these measurements are reported eliminates any confusion or wasted effort regarding data collection. Clarity regarding the data facilitates the negotiation of penalties for non-performance.

Routing Considerations

When implementing an L3 MPLS VPN service, it is important to understand whether any changes are needed to the routing protocol used by an enterprise customer, how this protocol interacts with the SP, and other routing issues. This section describes some general issues and includes the following topics:

- Route Limits, page 15
- Routing Protocol Support and Behavior, page 15
- Backdoor Connectivity Options, page 15
- Routing Convergence, page 16
- Load Balancing, page 16

Specific details and considerations to keep in mind when implementing the L3 MPLS VPN are described later in this document.

Route Limits

SPs may impose limits on the number of routes that can be advertised by the customer. It is important to understand what these limits are and what notifications, warnings, or repercussions occur if the limits are exceeded.

If route limits are imposed, take careful note of any summarization that may be broken when the network is transitioned to the SP L3 MPLS VPN service. This is especially important in the case of hub-and-spoke enterprise designs where the hubs summarize the spoke address assignments. When the spoke sites transition to a Layer 3 MPLS VPN service, this summarization may break and the number of entries in the enterprise routing table may increase, depending on the original level of summarization.

Routing Protocol Support and Behavior

Because a Layer 3 MPLS VPN service offering interacts with the SP at Layer 3, some routing environment considerations must usually be taken into account. This occurs when using a routing protocol on the PE-CE link that is different from the IGP used in the current enterprise environment. For example, an enterprise might use EIGRP as their IGP and eBGP as the PE-CE protocol. In this scenario, there must be careful consideration of administrative distance, redistribution between EIGRP to/from eBGP, and routing loops that might occur.

Backdoor Connectivity Options

When backdoor connectivity (connectivity outside the MPLS VPN service) is used, there is potential for problems such as routing loops or sub-optimal routing. Depending on the protocol being used on the PE-CE link, various methods for implementing backdoors are available, but you need to understand what is supported by the SP. For example, are OSPF Sham Links supported? Does the PE support BGP cost community or Site of Origin (SoO)?

Routing Convergence

Because the SP in a Layer 3 MPLS VPN service is participating in routing with the enterprise, routing convergence depends on the SP network routing convergence. Some SPs do not provide a convergence SLA, but you should still understand the approximate convergence times for failures such as PE-CE link failure or CE route withdrawal. You should find out whether there is any flexibility in adjusting convergence times, and ensure that they are acceptable for your application needs.

Load Balancing

When a site (CE) is connecting to multiple PEs, it makes sense to use all the links. CE-PE load balancing is controlled by the enterprise. PE-CE load balancing is controlled by the SP, so you should find out whether the SP supports this.

BGP multipath features employed in the SP environment let you load balance PE-CE traffic. Such load balancing lets the PE router forward against multiple BGP routes for the same customer prefixes, assuming they meet the BGP multipath requirements. This feature allows for load balancing across multiple BGP paths, but at the loss of determinism regarding the path traffic takes for a specific destination. For further information, see the following URL:
http://www.cisco.com/en/US/products/sw/iosswrel/ps1839/products_feature_guide09186a00800b5d5f.html

Without this load balancing feature, BGP normally selects a single best path, which may overload traffic on one link. One way to avoid this requires you to decide the prefixes that are preferred over each link in a multihomed environment. This solution requires the high administrative overhead of specifying prefixes, attribute setting, and so forth, but provides deterministic traffic flow.

Multihop eBGP can also be a useful load balancing tool. When multiple links exist between the CE and PE, eBGP can be configured between the loopbacks of the PE and CE routers. For more information, see the following URL:
http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a00800945bf.shtml.

Layer 2 Access to the MPLS VPN Service

Access to the MPLS VPN network is provided over a link between the CE and PE routers. Service providers usually offer a wide range of connectivity options, such as Frame Relay, ATM, and Ethernet. This section describes some of the Layer 2 access options and includes the following topics:

- Support of Existing Layer 2 Capabilities, page 16
- Access Speed Range, page 17
- Link Failure Detection, page 17

Support of Existing Layer 2 Capabilities

You should consider the existing Layer 2 capabilities available at various sites, and whether the provider can offer connection options to match these capabilities. Otherwise, the cost of establishing Layer 2 connectivity to these “unmatched areas” should be considered.

Access Speed Range

You should also consider the range of access speeds supported in each access method, and whether the purchased access speed is less than the access method speed. For example, you might purchase a 30 Mbps rate on a 100 Mbps Fast Ethernet access method. In this case, some SPs perform traffic policing to enforce the purchase rate. This requires the CE to perform shaping to avoid overrunning the policer configured on the PE. In a managed CE environment, traffic shaping is configured by the SP.

Link Failure Detection

It is important to understand the link failure detection mechanisms provided by the access methods used in a specific deployment. Access methods may have an inherent Layer 2 keepalive mechanism that supports link failure detection. Some access methods, such as Ethernet, may not appear down in the event of a failure on one end, which makes it difficult to detect failure. This depends on the physical configuration and the available features, such as Bidirectional Forwarding Detection (BFD).

QoS Capabilities

The support for end-to-end QoS provided by MPLS helps ensure that critical networking traffic is given the appropriate priority through the network. You should discuss your requirements related to the types of traffic that need specific priorities.

It is important to understand the classes of service (CoS) that are available in the SP network. Can CoS values sent from the CE to the provider network be preserved until they reach the remote CE? If not, is it possible to map the CoS values used by the customer to the CoS values used by the SP so that they can be mapped back to the customer values at the opposite end of the VPN?

As mentioned earlier, providers may partner with other providers to interconnect MPLS VPN networks to provide global services, and these partnerships may affect QoS. Assignment of CoS values may differ from one provider to another, making it necessary to translate CoS values between providers. This is something that is typically made possible by the agreement between the MPLS VPN providers. This agreement must specify CoS equivalencies. You should understand these values and equivalent values to ensure that the SP QoS capabilities are sufficiently transparent to support your requirements.

Multicast Capabilities

Multicast allows information to be efficiently distributed between a single multicast source and many receivers. Multicast has many uses, including financial applications, software downloads, and audio and video streaming.

Initially, MPLS VPNs did not support IP multicast traffic. In early deployments, support for multicast traffic was provided through GRE tunnels. GRE tunnels were built between CE routers, and all multicast traffic between VPN sites was encapsulated using GRE. However, in this scenario, optimal multicast routing requires a full mesh of GRE tunnels, which is not scalable or manageable with a large number of VPN sites.

Multicast VPN (mVPN) provides a more scalable method of transporting multicast traffic between VPN sites. The details of mVPN can be found in the *Multicast VPN Design Guide* at the following URL: http://www.cisco.com/en/US/tech/tk828/tech_digest09186a00801a64a3.html.

You should know whether the provider supports mVPN as part of their MPLS VPN services. If not, what alternative solutions do they provide for multicast?

If mVPN is supported, are data multicast distribution trees (data MDTs) used? If so, what is the threshold and how many data MDTs are configured for customer data streams? A data MDT is a group that is dynamically created when the customer multicast traffic stream exceeds a configured threshold on the PE. The purpose of the MDT is to restrict transmission of a stream to the remote PEs that are interested. These numbers are important because when the throughput of the customer stream surpasses the data MDT threshold and the maximum number of data MDTs already exists, the group addresses are reused. This may mean that some PEs receive CE data to which they have not subscribed.

Are Auto-RP and bootstrap router (BSR) supported? BSR messages should not be exchanged between different domains, because routers in one domain may elect rendezvous points (RPs) in the other domain, resulting in protocol malfunction or loss of isolation between the domains.

Security

MPLS VPN networks provide the same level of security as L2 VPN networks such as Frame Relay and ATM. MPLS VPN networks offer address space separation, routing separation, and are resistant to attacks and label spoofing. In an MPLS environment, a VPN customer may perform IP source address spoofing, but because there is a strict separation between VPNs and between the VPN and the core, this type of spoofing remains within the VPN where it originated. However, because MPLS VPN networks are part of a shared infrastructure, there are security considerations when evaluating an SP. This section describes some of these issues and includes the following topics:

- Shared Infrastructure, page 18
- MPLS Core Protection, page 18
- Other Security Policies, page 18

Shared Infrastructure

Are Internet and VPN access provided over the same core infrastructure? It is helpful to understand the security measures in place to avoid having one network service affecting the other.

A VPN-only service is more secure because there is no chance of attacks from the Internet. However, the level of risk associated with a shared core infrastructure is acceptable for most customers. The SP may offer separate PE routers for Internet and VPN access. However, this usually comes at a higher cost to the customer.

MPLS Core Protection

It is important to MPLS VPN customers that the SP core is protected from outside attacks. This prevents attackers from using the SP core to attack the VPNs. You can ask the SP to disclose information about the security of their infrastructure when evaluating the SP.

Other Security Policies

What policies are in place to prevent deliberate or accidental misconfiguration within the SP that may expose the customer VPN to attacks from the Internet or other VPNs? MPLS VPNs are as secure as L2 VPNs, but people make mistakes. It is important that the proper policies are in place to mitigate the risks.

Connecting to an MPLS/VPN Service Provider

If choosing a managed CE service, the task of connecting to the service is placed on the service provider. However, you should understand the necessary considerations because many of them involve you. This section includes the following topics:

- Migration
- CE-PE Routing Considerations, page 24
- Default Route Handling, page 46
- Load Balancing, page 53
- Multihoming Scenarios, page 54
- Quality of Service Considerations, page 60

Migration

Migrating from a traditional Layer 2-based VPN such as Frame Relay or ATM, where the service provider does not participate in Layer 3 routing with the enterprise, to a Layer 3 MPLS VPN service, where the provider does participate in Layer 3 routing with the enterprise, offers several challenges. IT managers who want to take advantage of the benefits of Layer 3 MPLS VPNs are looking for guidance on how to plan for these challenges and how to manage the migration as transparently as possible. This section addresses some of the steps that should be taken to assist in a smooth migration. Because every enterprise migration case is different, the examples in this section emphasize some of the more important things to consider to help facilitate your own migration.

Assessing Existing Network Elements and Enterprise Requirements

When you are considering migration, it is assumed that you have completed the task of actually choosing a service provider that meets the needs of your enterprise. As part of making this decision, you now have a good idea of what services the provider offers. You can now focus on assessing your network and identifying some of the important elements and requirements of the network that determine the overall complexity of the migration, such as the following:

- Knowing your internal site routing protocols and your options of PE-CE routing protocols offered by your provider. If your current internal routing protocol(s) are different from what the provider allows on the PE-CE links, some form of redistribution is required, unless you have the option to change your internal routing protocols.
- Which enterprise sites will be multi-homed or need redundancy? Is load-balancing required for these sites? The enterprise may decide to multi-home some sites, while other sites might be single-homed only to the provider.
- Do any sites require backup services? Are these services provided by the service provider or by the enterprise?
- If backup needs are provided by the enterprise, maintaining part of the existing network may provide the backup capability. Identify the parts of the existing network that will be maintained.
- Identify temporary transit sites. The purpose of the transit site is to allow migrated sites to communicate with non-migrated sites. Typically, hub sites or sites that get the most traffic, such as data center sites, are selected as transit sites.

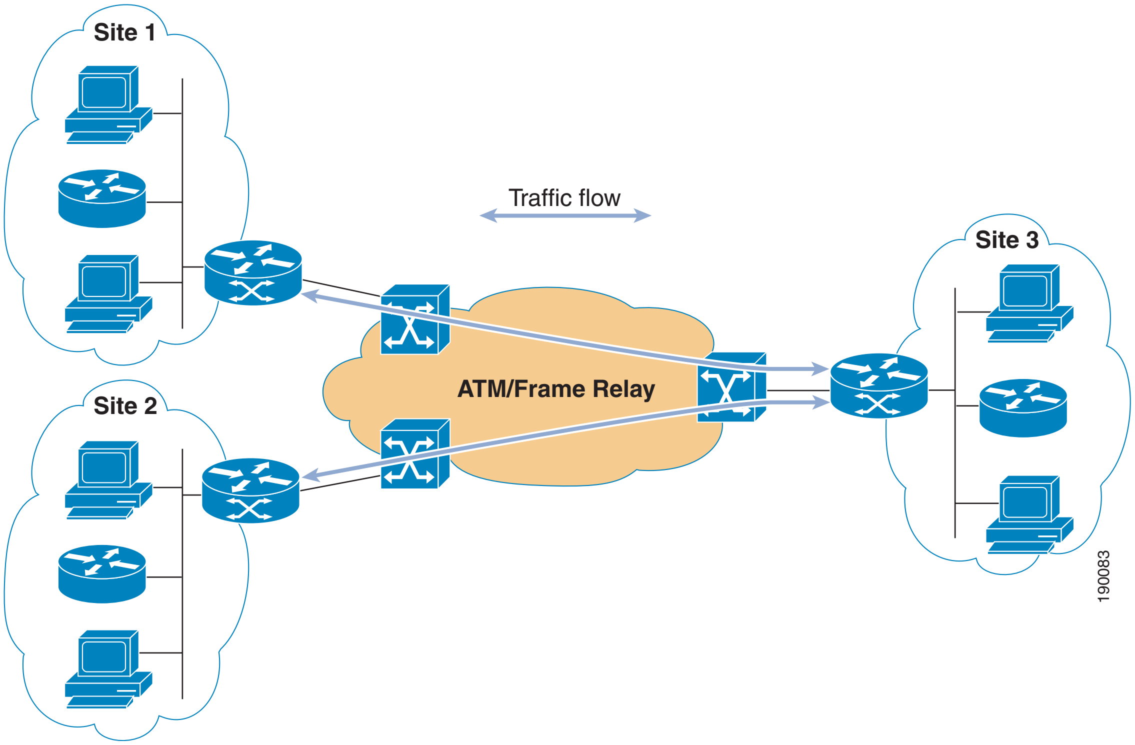
After the existing network elements and requirements have been assessed, you can compare your findings with what the service provider is currently offering to help plan the migration. For example, if you have determined that some sites require a high-speed backup solution but the service provider does not provide backup services, you must develop an in-house solution.

Physical Migration

After the existing network elements have been assessed, you can start looking at the physical migration from the existing WAN core to the L3 MPLS VPN of the service providers, which is the first step of migration.

The physical migration starts with the site or one of the sites that have been designated as a transit site. In Figure 6, Site 3 is a hub site that has been selected as a transit site for traffic between Sites 1 and 2.

Figure 6 Selecting the Transit Site

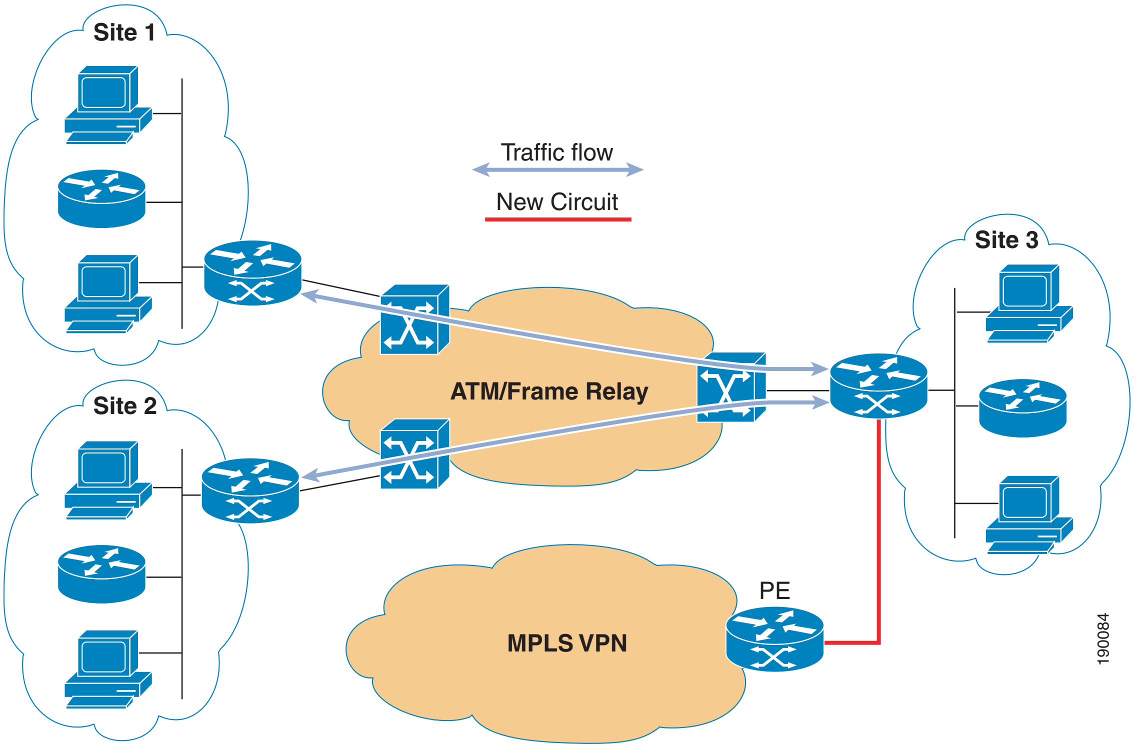


Start by provisioning a new circuit that attaches the CE at Site 3 to the PE in the MPLS network of the service provider. The original circuit into the original WAN stays intact at this time. Site 3 is the transit site, which means that it maintains connectivity between the sites that have migrated and those that have not. Because of this, the new PE-CE circuit that is provisioned must have enough bandwidth to accommodate the extra traffic.

After the new circuit has been provisioned for the new PE-CE link, Layer 3 connectivity over that link should be established. If a routing protocol is used over the PE-CE link, the RP peering can also be established at this time. The routing protocol to be run over this link is most likely determined by what the service provider supports. As previously mentioned, if this routing protocol is different from what is running internally to the site, some redistribution is required, and steps should be taken to avoid routing loops or sub-optimal routing that can occur with redistribution. PE-CE routing considerations are discussed subsequently in this document.

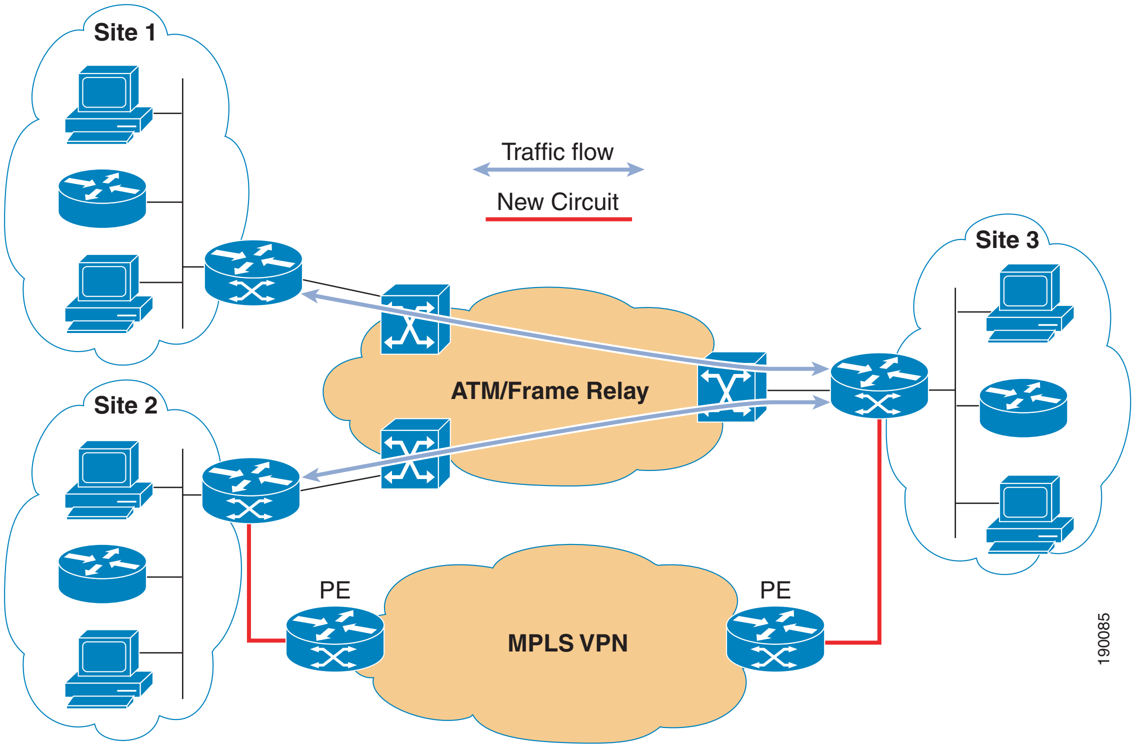
At this point, Layer 3 connectivity and RP peering have been established over the new circuit. Traffic between the sites continues to use the original WAN because no routing information is being exchanged over the MPLS network at this point. Figure 7 shows the state of the topology with the new circuit established.

Figure 7 New Circuit Established



You can now migrate another site. In this case, Site 2 is chosen. Again, a new circuit is provisioned that attaches the CE at Site 2 to the PE in the MPLS network of the service provider. The existing connection stays intact, as shown in Figure 8.

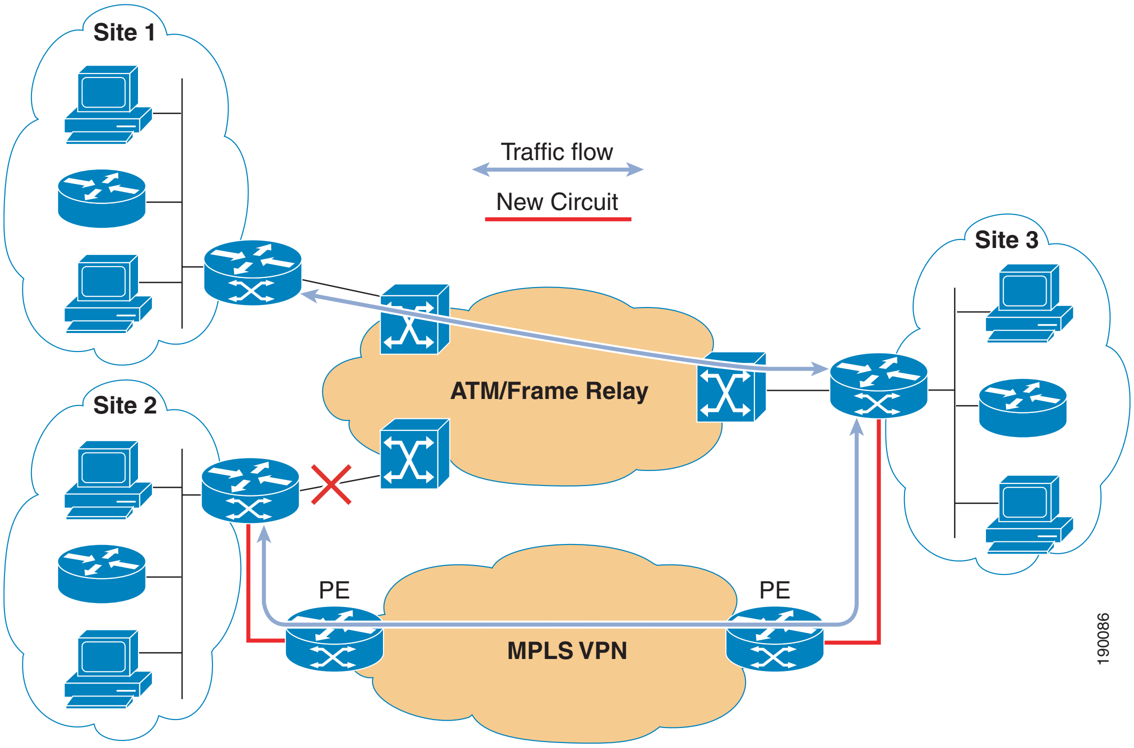
Figure 8 Migrating Site 2



After the physical circuit is in place, Layer 3 connectivity over the new link needs to be established. In most cases, the same routing protocol that is used over the Site 1 PE-CE link is also used on this link. Site 2 now has two connections to Site 3. One connection is via the new MPLS network, and a second is via the original WAN network. Routing information now gets exchanged over the MPLS network as well as over the original network. If at this point traffic is not flowing over the MPLS network, it may be necessary to influence the routing by manipulating some of the routing metrics so that the path over the MPLS network is the preferred path between Site 2 and Site 3. After the path through the MPLS network is established as the preferred path, the Site 2 connection to the original WAN can be disconnected.

At this point, you can see why Site 3 is called a transit site, as shown in Figure 9.

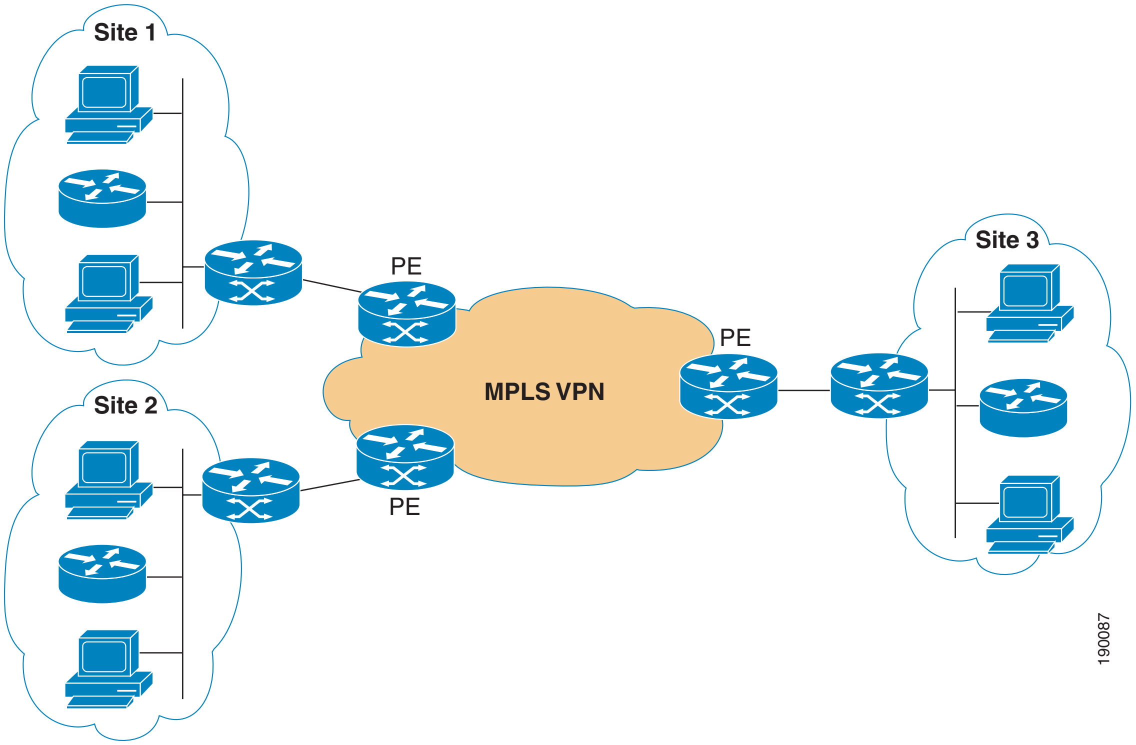
Figure 9 Using Site 3 as the Transit Site



In Figure 9, Site 1 is still communicating with Site 3 and Site 2 through the original WAN network, while Site 2 is now communicating with Site 3 and Site 1 through the new MPLS network.

Migrating Site 1 is your next step. Following the same procedure as before, first establish a new circuit to the MPLS network and then establish Layer 3 connectivity over the new circuit. After the new path over the MPLS network has been established as the preferred path, the original circuit can be disconnected, resulting in the network shown in Figure 10.

Figure 10 Original Circuit Now Disconnected



This site-by-site approach can be followed until all sites have been migrated.

Because CE-PE routing protocols are a crucial piece to Layer 3 MPLS/VPNs, the next section describes their behavior in this environment.

CE-PE Routing Considerations

This section includes the following topics:

- Using BGP for CE–PE Routing, page 24
- Using OSPF for CE-PE Routing, page 32
- Using EIGRP for CE-PE Routing, page 41

Using BGP for CE–PE Routing

BGP is one of the most common protocols used for routing between CE and PE devices. This section lists some important considerations to keep in mind when using BGP as the PE-CE protocol, and includes the following topics:

- BGP AS Allocation Schemes, page 25
- Using a Backdoor Link with BGP as the PE-CE Protocol, page 29
- Proper Filtering, page 31

BGP AS Allocation Schemes

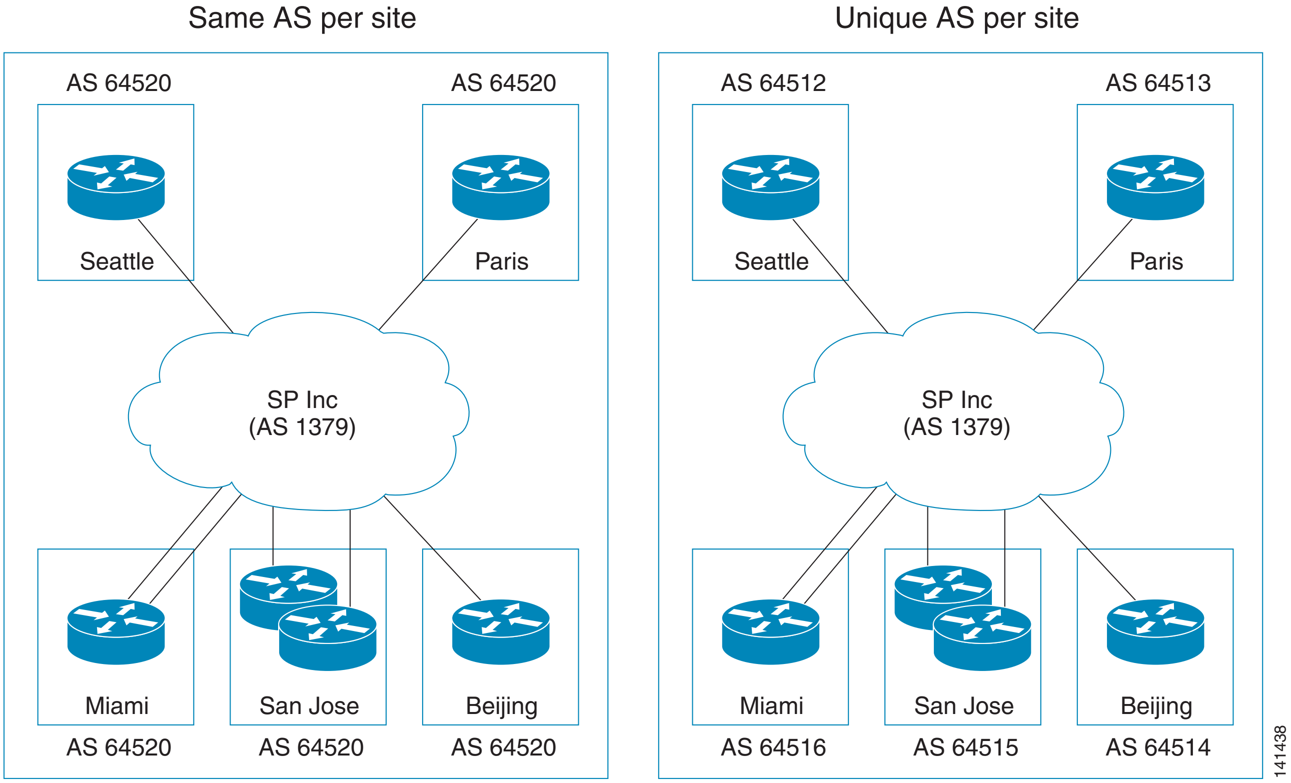
BGP requires that each BGP speaker be identified by an Autonomous System (AS) number. After choosing BGP as your PE-CE protocol, you must next determine the BGP AS allocation scheme. The selection of a BGP AS number for enterprise sites is an important consideration because it affects other aspects of network behavior, including load balancing, route-loop avoidance, and site characterization over the origin AS.

Most SPs offer two options for AS allocation:

- The same BGP AS for every customer site
- A unique BGP AS for each customer site

These options are illustrated in Figure 11. The left side shows an enterprise where every customer site uses AS 64520 to form an eBGP peering relationship with the SP, which uses AS 1379. The right panel illustrates an enterprise that allocates a unique AS for five sites, using the range 64512 through 64516.

Figure 11 BGP AS Allocation Schemes



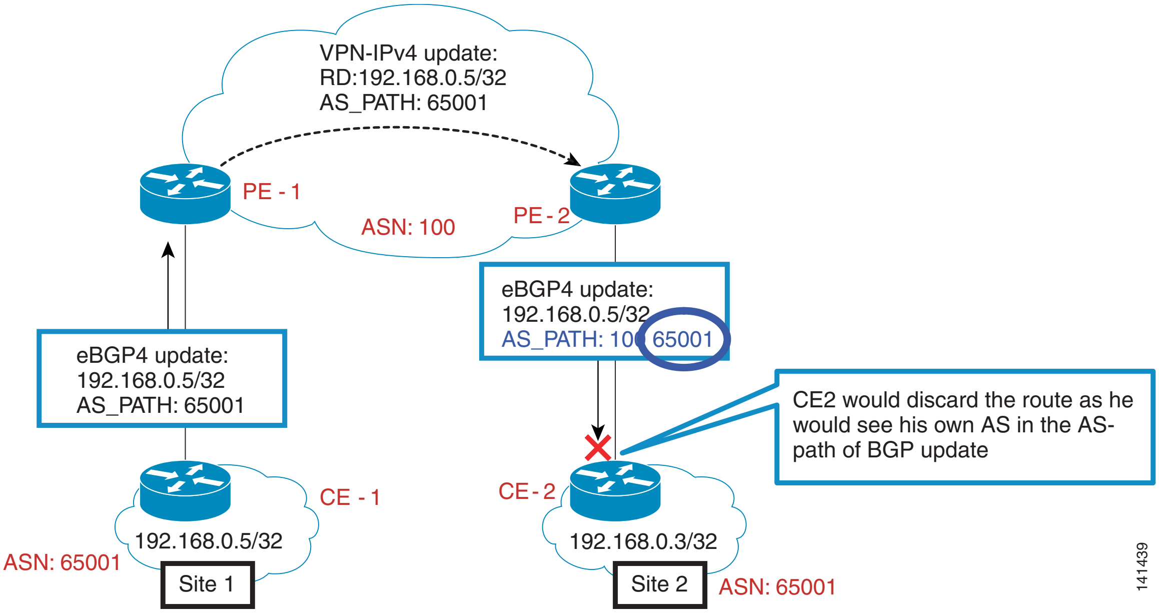
One of the main advantages of allocating a unique AS per site is that you identify the originator of the route by noting the origin BGP AS in the AS PATH attribute. This quick identification simplifies troubleshooting. Furthermore, easy origin identification allows simple AS-path filters to perform BGP route manipulation for a particular site.

However, a unique AS for each site limits the number of BGP speaking sites to the number of available BGP AS numbers. The available BGP range depends on the enterprise and the willingness of the SP to support public BGP AS numbers (1–64511). You should normally use the private BGP AS range (64512–65535) and never use BGP AS numbers unless they are registered to you. However, with an L3 MPLS VPN service, using unregistered AS numbers may not be a problem if the BGP MPLS VPN announcements are not injected into the public Internet routing table.

One of the advantages of using the same AS for every site is that it reduces the chance of AS collisions. However, the use of the same AS for every customer site also creates some complexity.

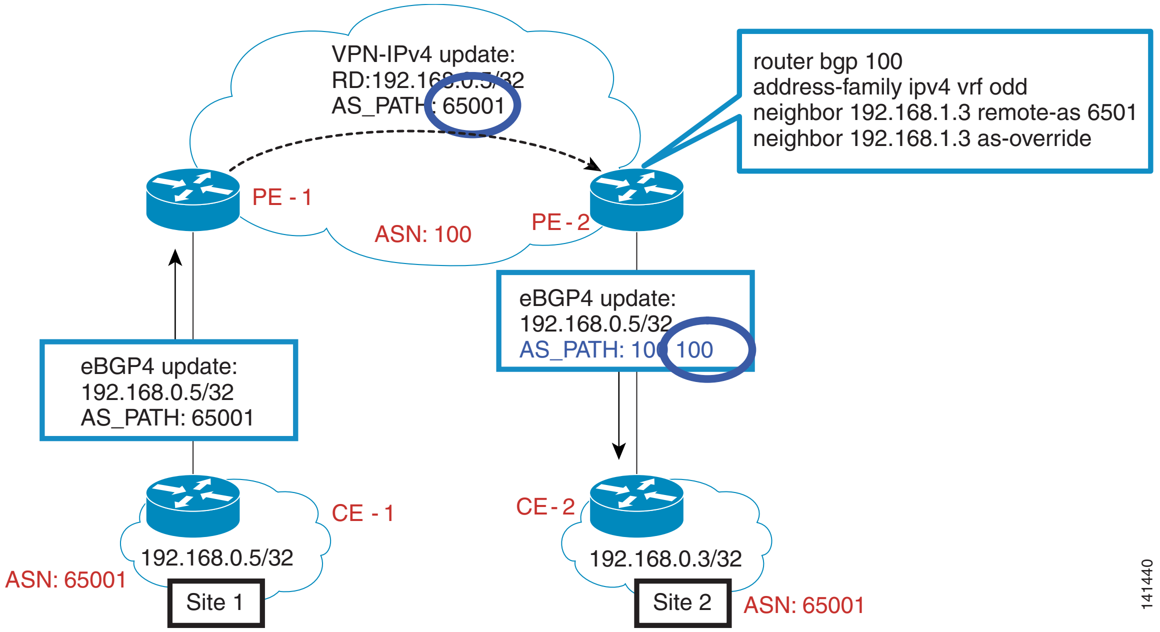
A BGP speaker performs AS loop prevention by verifying that the AS PATH contains its own AS number. This is illustrated in Figure 12, where Site 2 rejects a prefix originated from Site 1 because CE-2 recognizes its own AS (65001) in the AS PATH of the received route for 192.168.0.5/32.

Figure 12 BGP AS-PATH Loop Prevention



To use the same AS number for every customer site, AS loop prevention must be disabled. This is typically done by requesting that the SP adjust the AS PATH through the use of the **as-override** command. This is illustrated in Figure 13, where PE-2 is configured with **as-override**. PE-2 replaces the neighboring peer AS CE2 (65001) with its own AS (100), when it is detected anywhere along the AS-PATH attribute of the advertised BGP route.

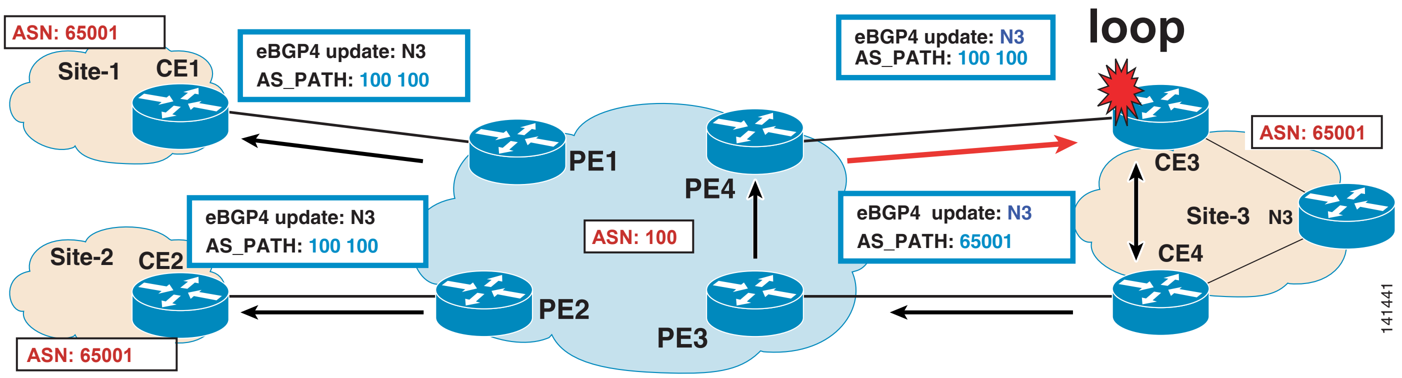
Figure 13 AS-Override Example



Mechanisms such as AS-override produce some additional complexity and configuration requirements on the SP. Another issue when using AS-override is that none of the BGP routes can be uniquely identified as originating from a specific site based on the AS-PATH. If the CE must identify the origin of the route based on some attribute, other mechanisms, such as BGP standard communities, should be considered. However, the latter option introduces additional configuration on the CE.

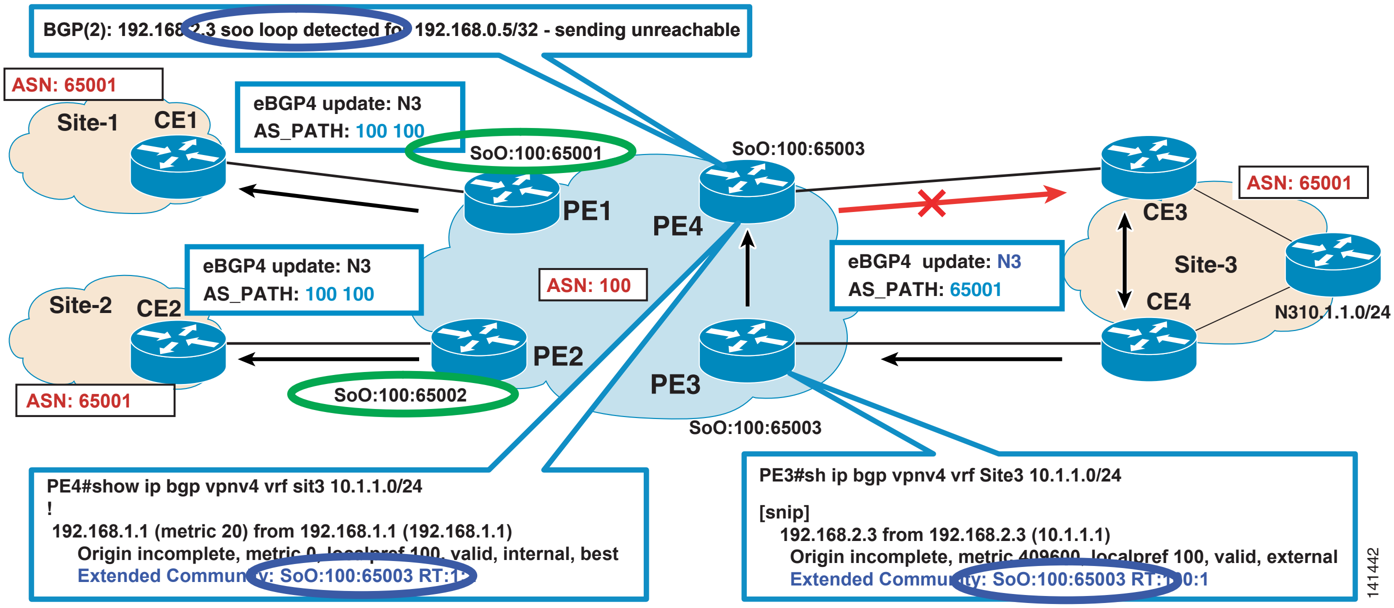
Rewriting the AS PATH essentially prevents the CE router from detecting a BGP loop, and can create problems in multihomed sites. Figure 14 illustrates a case where a route loop occurs. Site 3 originates the N3 route. CE4 advertises the N3 route to PE3, passes it to PE4, PE1, and PE2, which then advertise it to their respective CE routers. Unfortunately, in the case of CE3, the N3 route is received and accepted because the AS-PATH has been adjusted. This creates a route loop because the N3 route is advertised back into Site-3, which originated the route.

Figure 14 BGP Route Loop with AS-override



Site of Origin (SoO) can be used to avoid an AS-override induced route loop. SoO is an extended community attribute attached to a BGP route used to identify the origin of the route. If the attached SoO is equal to the configured SoO for a BGP peering, the route is blocked from being advertised; thereby avoiding a route loop. An example of SoO is shown in Figure 15.

Figure 15 Site of Origin (SoO) Example



```
BGP(2): 192.168.2.3 soo loop detected for 192.168.0.5/32 - sending unreachable
```

```
PE4#show ip bgp vpnv4 vrf sit3 10.1.1.0/24
!
 192.168.1.1 (metric 20) from 192.168.1.1 (192.168.1.1)
    Origin incomplete, metric 0, localpref 100, valid, internal, best
    Extended Community: SoO:100:65003 RT:1
```

```
PE3#sh ip bgp vpnv4 vrf Site3 10.1.1.0/24

[snip]
   192.168.2.3 from 192.168.2.3 (10.1.1.1)
     Origin incomplete, metric 409600, localpref 100, valid, external
     Extended Community: SoO:100:65003 RT:100:1
```

The PE4-CE3 and PE3-CE4 BGP peerings are configured with a SoO value of 100:65003. This configuration performs the following logic:

1. Any BGP advertisement received from these neighbors has an attached SoO equal to the configured value.
2. A check is performed on any BGP advertisement to these neighbors so that a loop is detected and the advertisement is blocked if the configured SoO value equals the attached SoO value.

The BGP route for N3 is received and the SoO value is attached to this route. PE3 propagates this route to PE1, PE2, and PE4. Note that the configured SoO for the PE1-CE1 and PE2-CE2 neighbor relationships are configured with SoO values of 100:65001 and 100:65002 respectively. Both PE1 and PE2 still advertise the N3 BGP route to their respective CEs because the configured SoO values do not match the attached SoO on the N3 route (100:65003). However, PE4 does not advertise the route to CE3 because the configured SoO for the PE4-CE3 neighbor relationship (100:65003) is equal to the attached value for the BGP N3 route.

The advantages and disadvantages of the various AS allocation methods are summarized in Table 4.

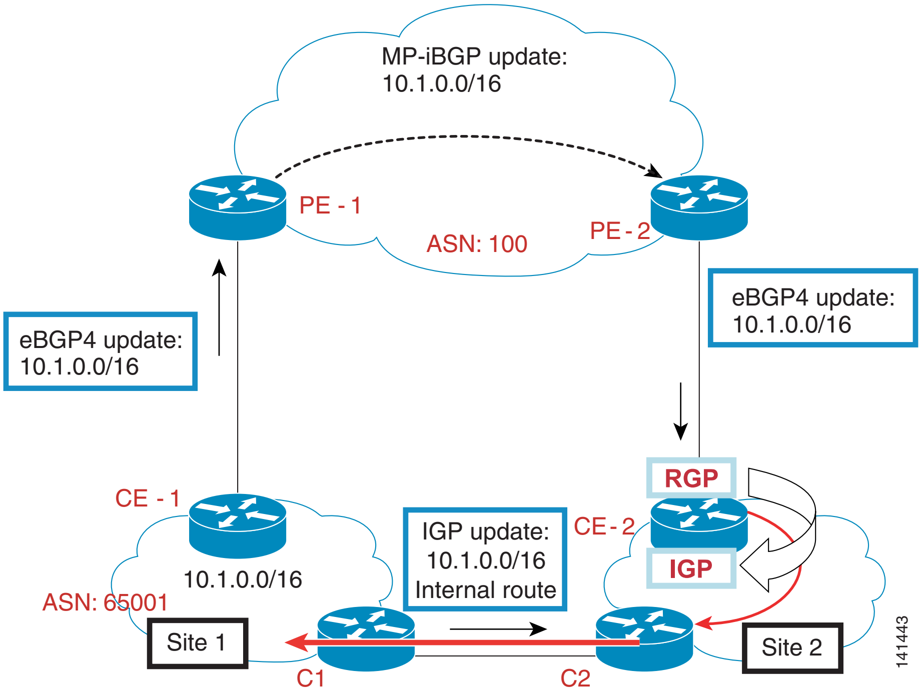
Table 4 Advantages and Disadvantages of AS Allocation Methods

BGP Allocation Method	Advantage	Disadvantage
Unique AS per site	Allows simple identification of route originator through the origin AS in AS-PATH attribute	Limits the number of customer sites to the number of available BGP AS May require allocation of AS numbers outside of private AS range (64512–65535) Requires more careful tracking of BGP AS allocation to avoid AS collision
Same AS per site	Reduces likelihood of AS collision when multiple providers are used	No site unique characteristic can be inferred from the AS-PATH Requires SP to rewrite AS-path via the use of AS-override (or customer configuration of allow-as) The use of AS-override or other mechanism essentially disables BGP AS loop prevention check, so alternate loop prevention mechanisms must be employed, such as SoO

Using a Backdoor Link with BGP as the PE-CE Protocol

This section describes how a backdoor link between customer sites is used, and the implications when implementing an L3 MPLS VPN. Figure 16 illustrates this topology.

Figure 16 Using a Backdoor Link with BGP

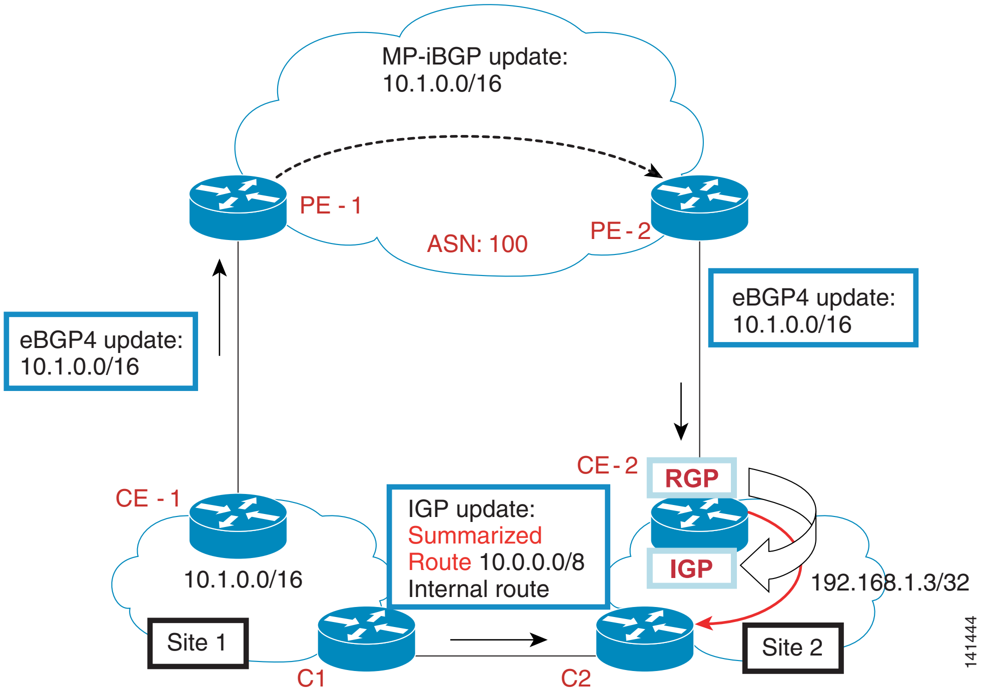


In this topology, two customer sites are connected to an MPLS VPN cloud. Each of the sites is running its own IGP. BGP is the PE-CE routing protocol. A network is being advertised from Site 1. PE1 receives this network from eBGP and in turn advertises it to PE2 through MP-iBGP. CE2 then receives it through eBGP from PE2. 10.1.0.0/16 is then redistributed into the Site 2 IGP as an external route.

At the same time, router C2 is receiving 10.1.0.0/16 from C1 through IGP. This routing update is an internal route. Now Site2 has two routes for 10.1.0.0/16: an external route from CE2, and an internal route from C2. Therefore, traffic in Site 2 destined for 10.1.0.0/16 uses the backdoor link because the internal route is preferred over an external route. This is not the desired behavior.

The backdoor link may exist from a legacy infrastructure and can be removed to solve the problem. But in many cases, backdoor links exist to provide redundancy and cannot be removed. One way to solve the problem is to summarize the routes on the backdoor link as shown in Figure 17.

Figure 17 Summarizing Routes on a Backdoor Link



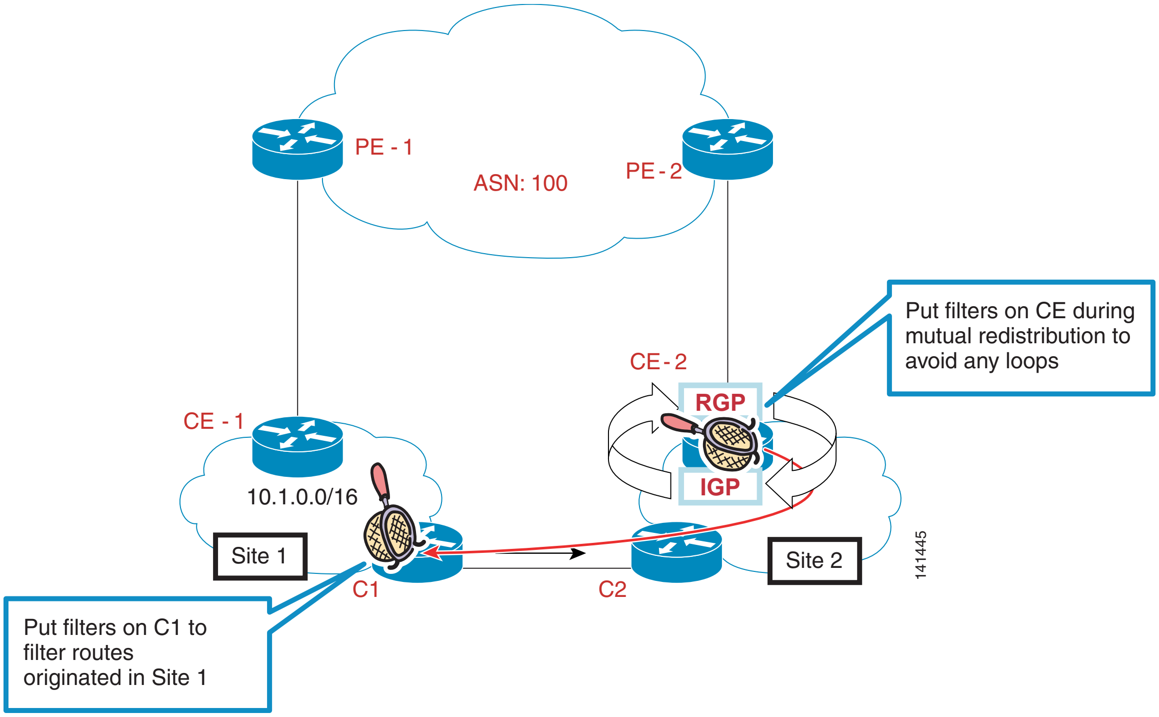
In this example, the external route for network 10.1.0.0/16 is still received from CE2. However, now on C1, the route is summarized to 10.0.0.0/8, and C1 receives the summarized route. Now traffic in Site 2 destined for 10.1.0.0/16 uses CE2 as the exit router because a more specific route is being received from CE2. The backdoor link is used only if the more specific route is lost.

Note Summarization requires special configuration when using OSPF. For OSPF, summarization is possible only on area border routers (ABRs). Therefore, to summarize you need to make C1 and C2 into ABRs. This means that the backdoor link is in area 0 while Site 1 and Site 2 are in a non-zero area or vice versa. Another possible solution for OSPF is running a different routing protocol between C1 and C2, and doing summarization while redistributing.

Proper Filtering

When a customer site is running an IGP and BGP is used as the PE-CE protocol, mutual redistribution must be done on the CE between the IGP and BGP. This can cause routes to get redistributed back into BGP, potentially creating routing loops. It is therefore recommended to use proper filters during mutual redistribution. Filters should be configured so that only site-specific routes are allowed to get redistributed into BGP, as shown in Figure 18.

Figure 18 Using Filters to Prevent Loops



Put filters on CE during mutual redistribution to avoid any loops

Put filters on C1 to filter routes originated in Site 1

Similarly, if a backdoor link exists as shown in Figure 18, there is a chance for routes originated in Site 1 to be learned back from C2. Therefore, you need to put filters on C1 and C2 to filter routes originated within the respective sites.

Using OSPF for CE-PE Routing

OSPF has been used as an IGP for a long time. This section discusses what you should consider when using OSPF as a PE-CE routing protocol.

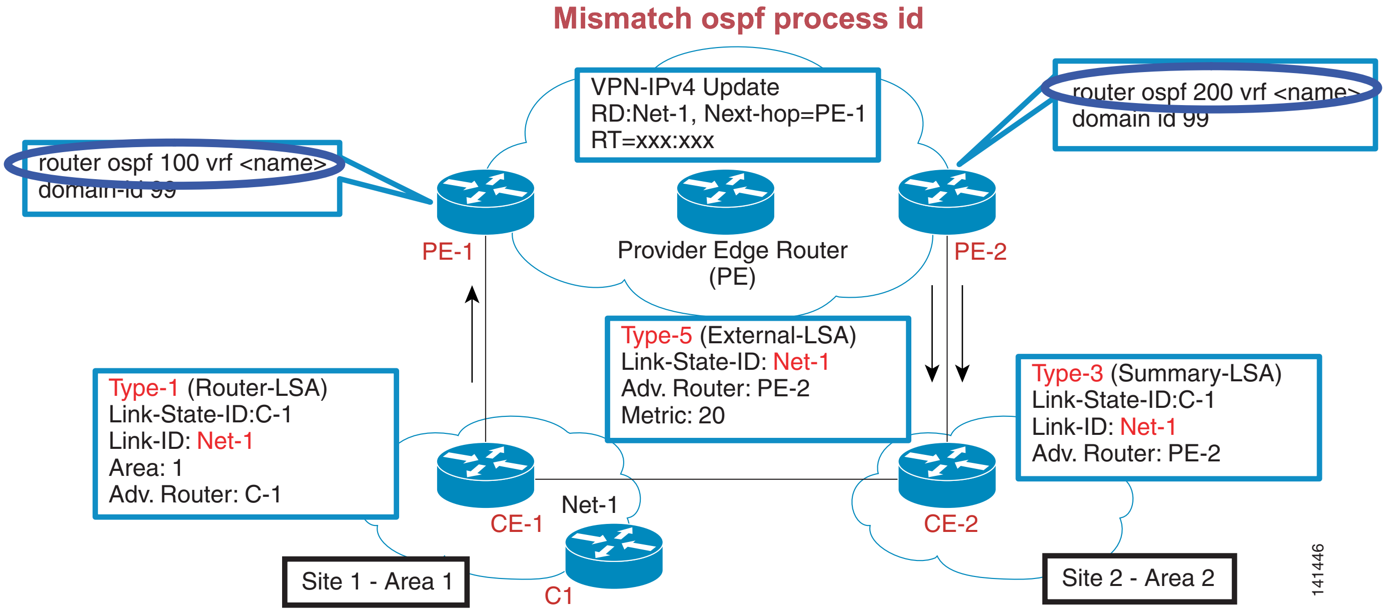
- Different OSPF Processes at Each Site, page 32
- OSPF Route Summarization Techniques Used with MPLS VPNs, page 33
- OSPF Area Placement Considerations, page 37

Different OSPF Processes at Each Site

In MPLS VPN networks, the OSPF process ID should match. Otherwise, external Type 5 routes are generated. In Figure 19, two organizations have merged. In this scenario, Site 2 expects Type 3 inter-area routes from Site1 from PE2 but instead receives external Type 5 routes. This happens because the OSPF process ID is different on the two sites.

When implementing an L3 MPLS VPN, the SP cloud appears as a single router from the OSPF perspective. Instead of removing and reconfiguring the OSPF process, the SP may configure the same domain ID on both ingress and egress PEs to solve the problem.

Figure 19 Sites with Different OSPF Processes



When Net-1 is advertised from CE-1 to PE1 as an OSPF LSA Type 1, the PE1 router converts it into an MP-iBGP update and advertises this update to PE-2. PE-2 converts this to LSA Type 5 when it sees that the OSPF process ID of the destination is different. In this scenario, CE-2 receives an OSPF external route and this should not happen. If you configure the same domain ID on both PE-1 and PE-2 under the OSPF configuration, this problem can be solved without any further OSPF configuration changes. After making this change, CE-2 receives Net-1 as an inter-area OSPF route.

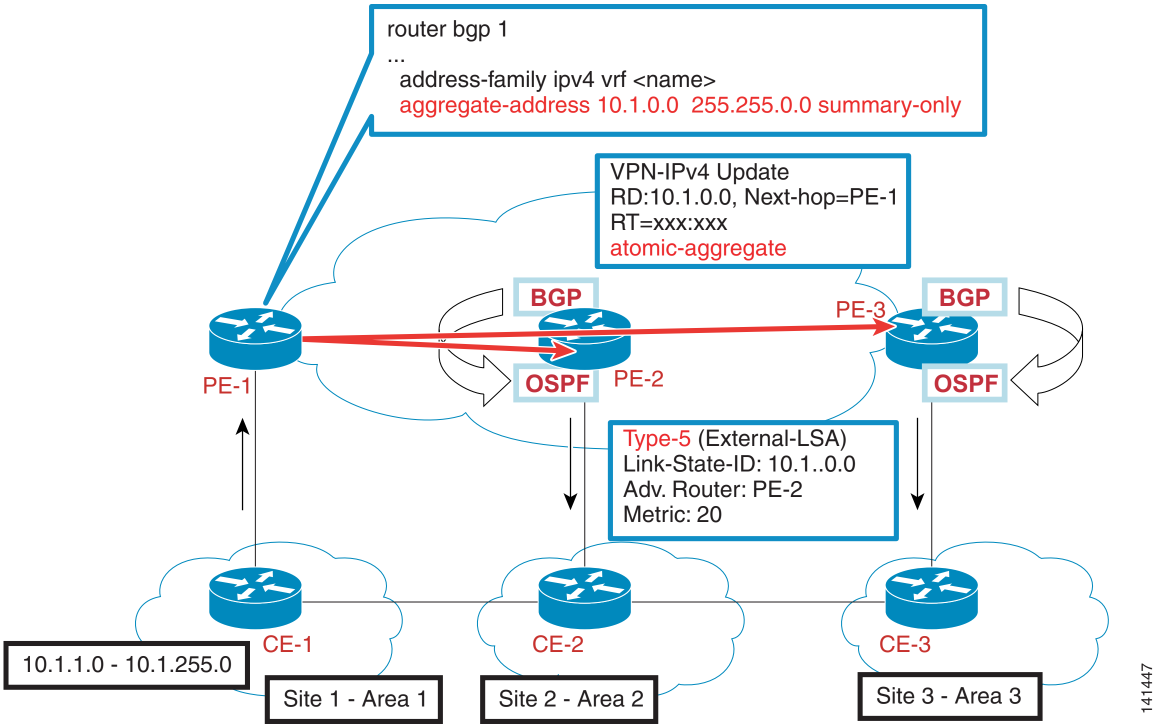
OSPF Route Summarization Techniques Used with MPLS VPNs

This section describes two types of OSPF summarization: ingress-side summarization and egress-side summarization.

Ingress PE-Based Summarization

If an MPLS VPN customer running OSPF as a PE-CE protocol wants to send a summary route to all other sites, it cannot be done because there is no ABR at the site. In this case, the PE router connected to this site can summarize in BGP and advertise the aggregate to all other sites. This is shown in Figure 20.

Figure 20 Ingress PE-Based Summarization



```
router bgp 1
...
 address-family ipv4 vrf <name>
 aggregate-address 10.1.0.0  255.255.0.0 summary-only
```

VPN-IPv4 Update
RD:10.1.0.0, Next-hop=PE-1
RT=xxx:xxx
atomic-aggregate

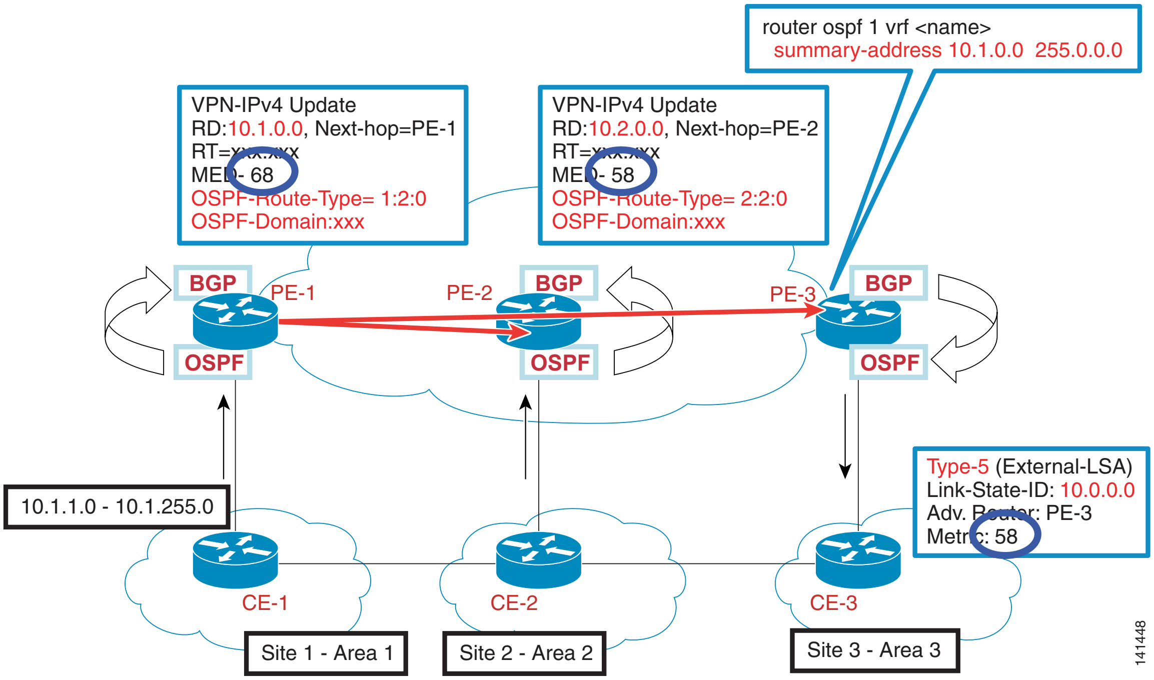
Type-5 (External-LSA)
Link-State-ID: 10.1..0.0
Adv. Router: PE-2
Metric: 20

CE-1 wants to send a summary route for 10.1.1.0 through 10.1.255.0 as 10.1/16 to all other sites from Site1/Area 1. PE-1 can summarize this address space in BGP and advertise an aggregate block to all other sites. CE routers at other customer sites see the aggregate as an external OSPF route.

Egress PE-Based Summarization

If an MPLS VPN customer running OSPF as a PE-CE protocol wants to send a summary route to only one or few sites, this cannot be done because there is no ABR at the site. In this case, the egress PE router connected to this destination site can summarize in BGP and advertise the aggregate to that site. This is shown in Figure 21.

Figure 21 Egress PE-Based Summarization

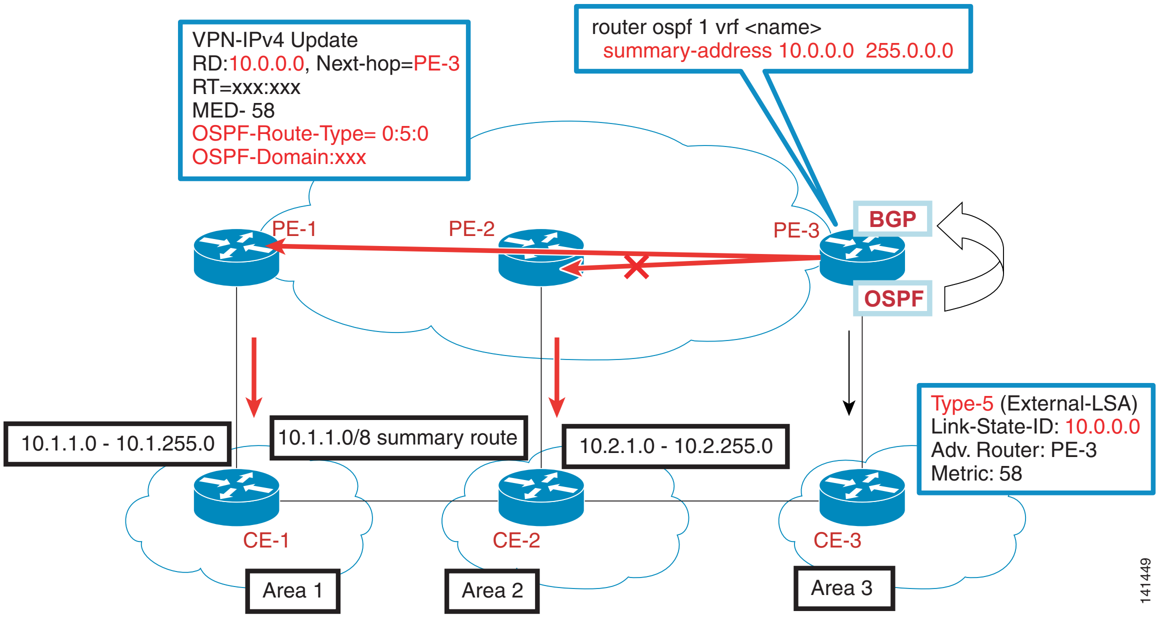


The customer needs to send a summary route to Site3-Area3. The customer cannot summarize from each individual site because there is no ABR within the sites at Area 1 or Area 2. In this scenario, the customer can ask the SP to summarize routes on PE3 for routes destined to Site3.

Loop Scenario

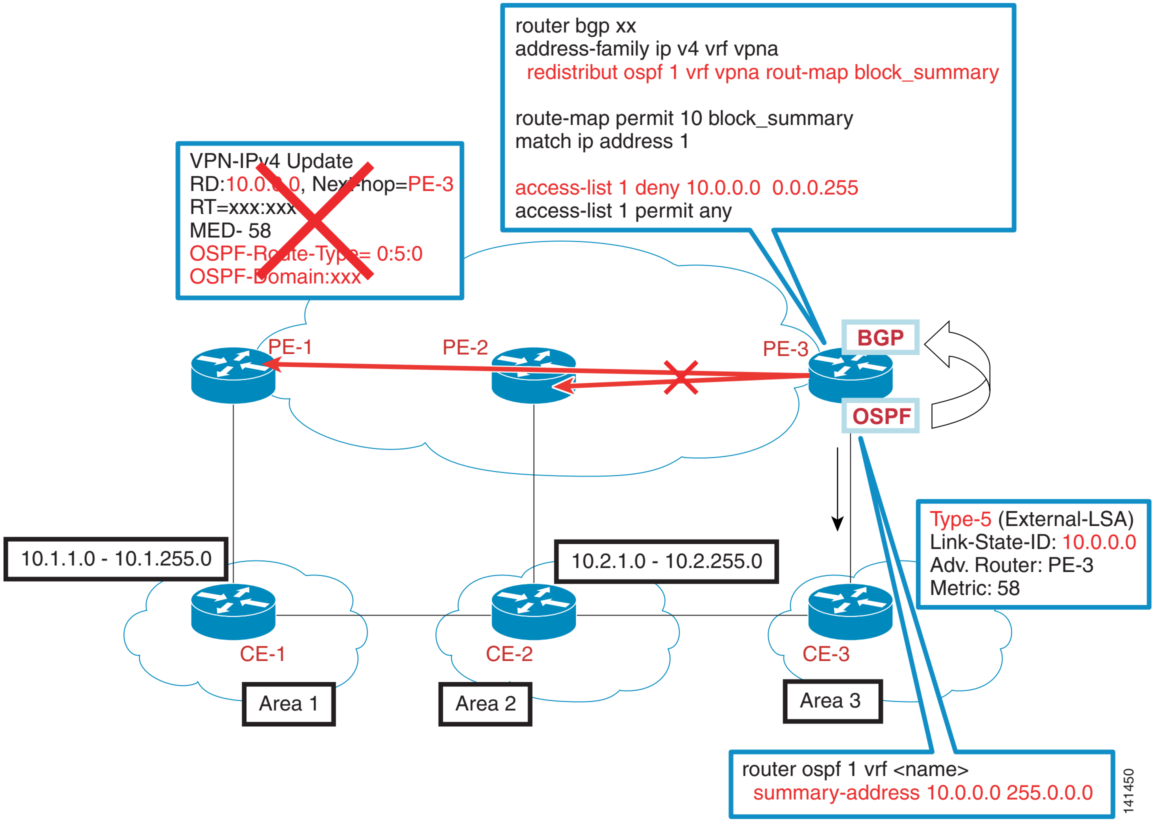
In another case, shown in Figure 22, the summary may originate in OSPF. The summary route 10/8 is propagated to all customer sites as a result of redistribution from OSPF into BGP. This can result in sub-optimal routing or routing loops.

Figure 22 OSPF Route Summarization May Create a Routing Loop



To prevent this situation, the summary route should be filtered while redistributing OSPF into BGP on PE3, unless it is desirable to send the summary to selected PEs. This can be done by using a route map called “block_summary”. This solution is shown in Figure 23.

Figure 23 Using a Route Map to Prevent a Routing Loop



```
router bgp xx
address-family ip v4 vrf vpna
  redistribut ospf 1 vrf vpna rout-map block_summary

route-map permit 10 block_summary
match ip address 1

access-list 1 deny 10.0.0.0  0.0.0.255
access-list 1 permit any
```

```
VPN-IPv4 Update
RD:10.0.0.0, Next-hop=PE-3
RT=xxx:xxx
MED- 58
OSPF-Route-Type= 0:5:0
OSPF-Domain:xxx
```

```
Type-5 (External-LSA)
Link-State-ID: 10.0.0.0
Adv. Router: PE-3
Metric: 58
```

```
router ospf 1 vrf <name>
 summary-address 10.0.0.0 255.0.0.0
```

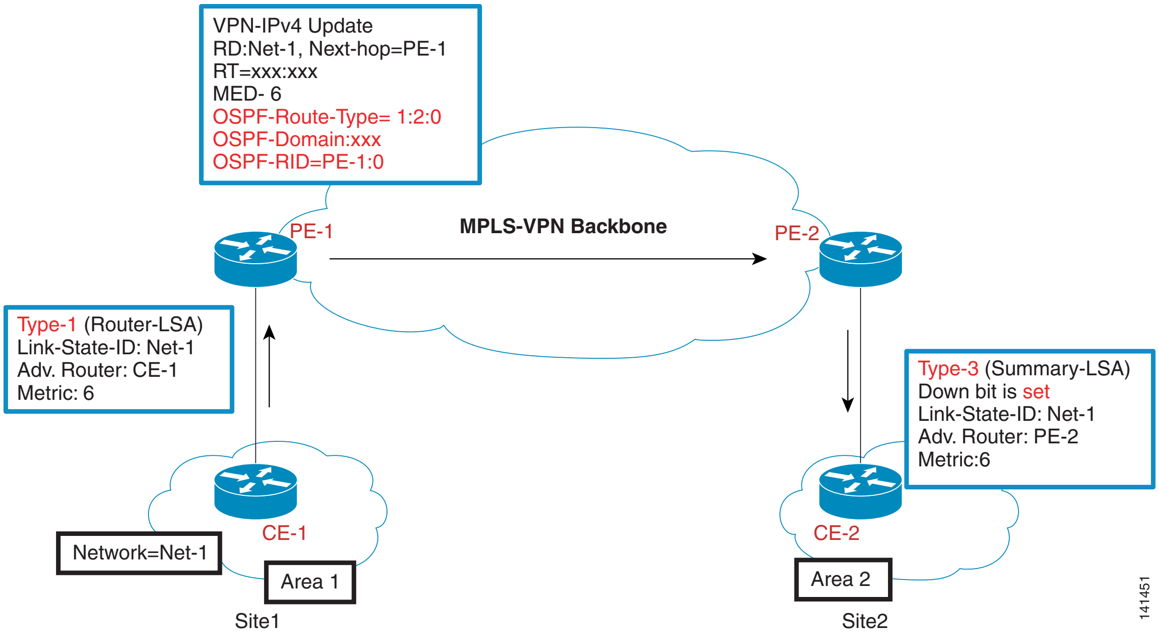
OSPF Area Placement Considerations

This section describes a few important concepts about the interaction of OSPF areas and the MPLS VPN backbone.

MPLS VPN Backbone Considered Area 0

Because the MPLS VPN backbone is considered Area 0, you do not necessarily need Area 0 at any site. Any Type 1 and Type 2 LSAs going across the MPLS VPN backbone are converted into Type 3 LSAs. Type 5 LSAs and external routes are received across the MPLS VPN backbone by the receiving OSPF routing process as Type 5 LSAs. This is shown in Figure 24, where a Type 1 LSA is converted to a Type 3 LSA as it goes across the MPLS VPN backbone.

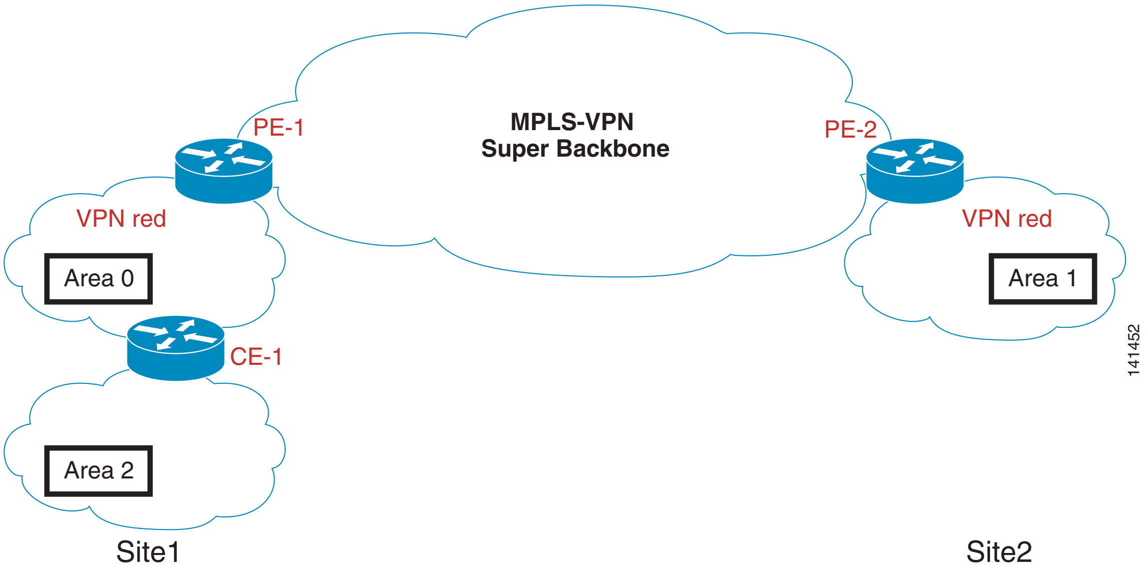
Figure 24 MPLS VPN Backbone Considered Area 0



Area 0 Adjacent to MPLS VPN

Area 0 must be adjacent to MPLS VPN or have a virtual link between Area 0 and the MPLS VPN backbone. The Area 0 site can be connected to the MPLS VPN backbone. However, if Area 0 exists, it must touch the MPLS VPN PE routers. Figure 25 shows this.

Figure 25 Area 0 Must Connect to the MPLS VPN Backbone

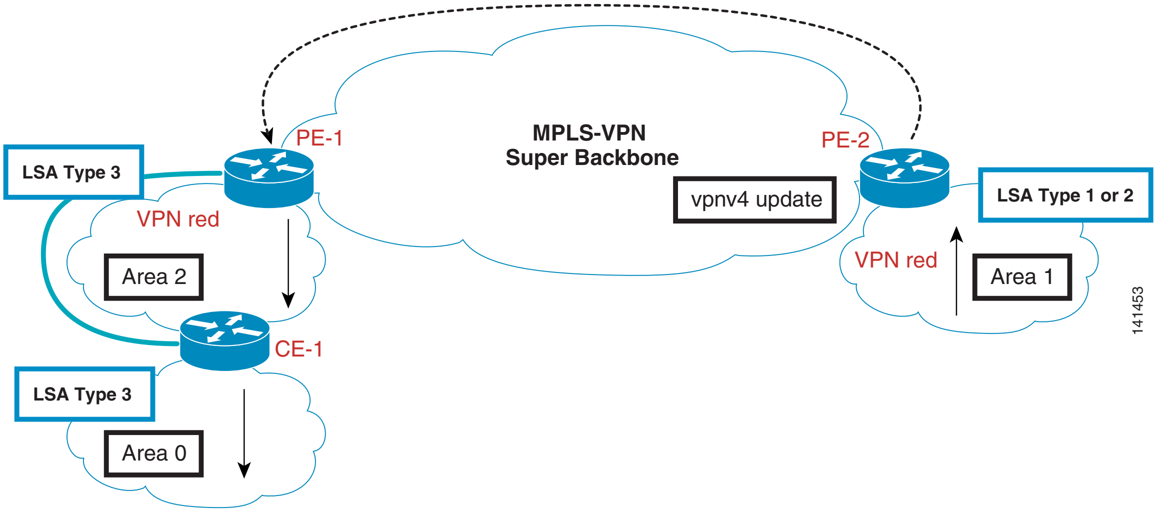


If Area 0 is not adjacent to the MPLS VPN backbone, you should set up a virtual link between Area 0 and the MPLS VPN backbone.

Note OSPF rule—Summary LSAs from non-zero areas are not injected into backbone Area 0. Therefore, inter-area routes do not appear unless a virtual link is created.

The scenario with a virtual link between Area 0 and the MPLS VPN backbone is shown in Figure 26.

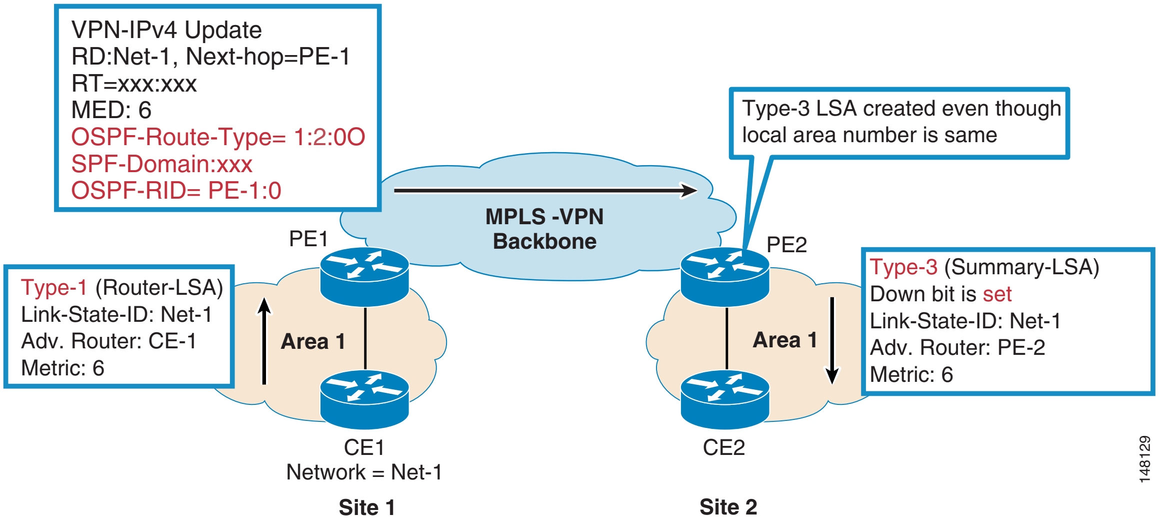
Figure 26 Virtual Link Between Area 0 and the MPLS VPN Backbone



Sites in the Same Area Without a Backdoor Link

In the scenario illustrated in Figure 27, the LSAs received at the sites are Type 3 LSAs (because any LSA transported across the MPLS VPN backbone are at least LSA Type 3), even though both sites are in the same area. If this is not desirable, you should consider using an OSPF Sham Link as shown in Figure 27.

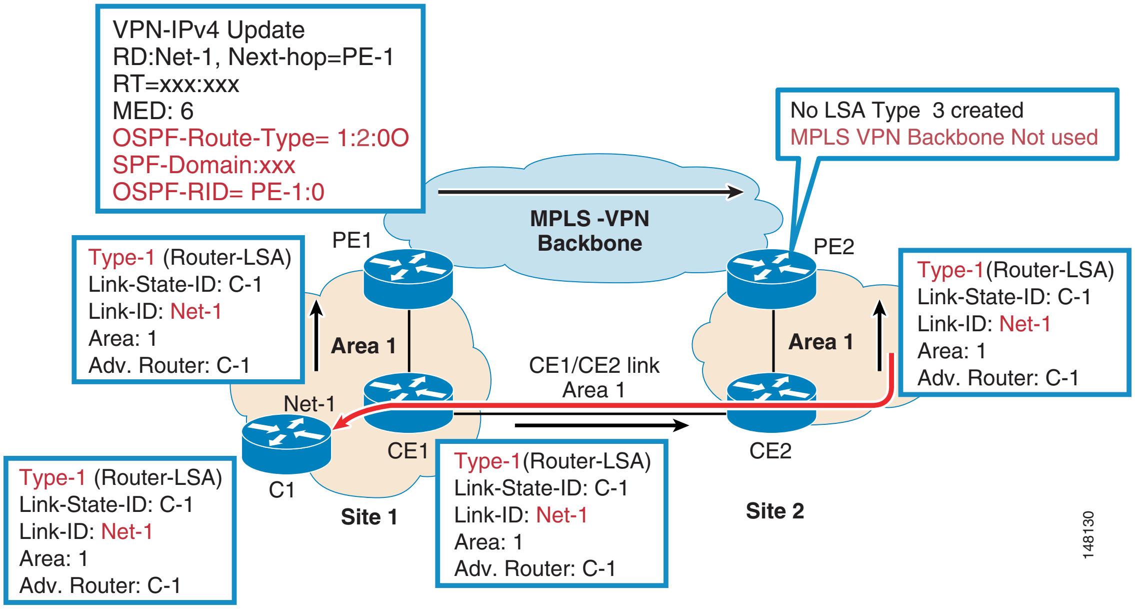
Figure 27 Using an OSPF Sham Link



Sites In the Same Area With a Backdoor Link

In Figure 28, the OSPF route is advertised to the MPLS VPN backbone. The same prefix is learned as the intra-area route over the backdoor link. PE2 does not generate Type 3 LSAs after a Type 1 LSA is received from the site. In this scenario, traffic is sent over the backdoor link instead of the MPLS VPN cloud.

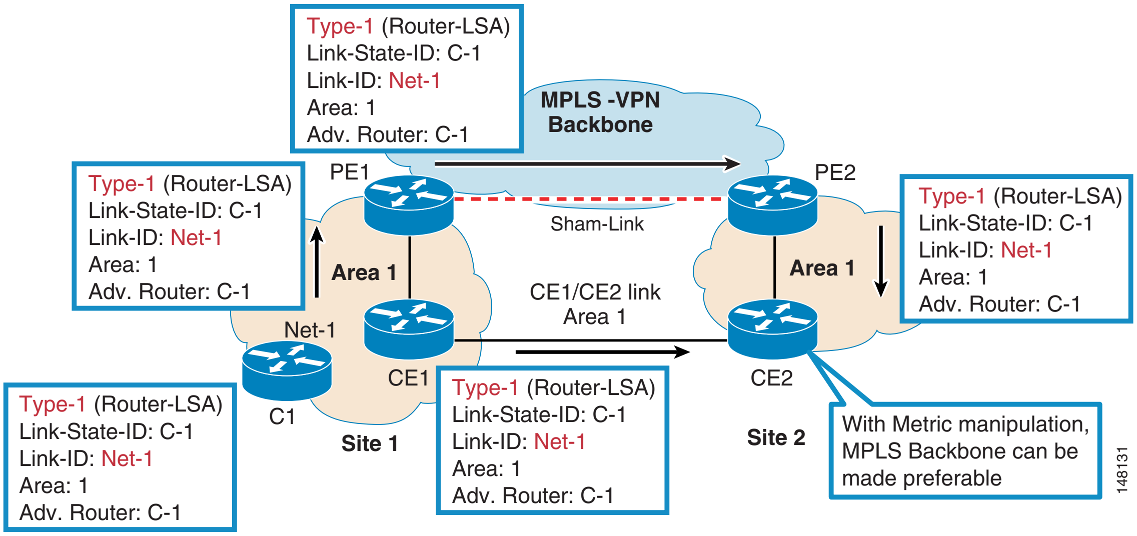
Figure 28 Sites in the Same Area with a Backdoor Link



Sites in the Same Area with Backdoor and Sham Links

A sham link is treated as a virtual link; it is a point-to-point and demand-circuit type link. OSPF adjacency can be established over a sham link. A sham link is reported in the router Type 1 LSAs originated by the two routers connecting to the sham link. Any Type 1 and Type 2 LSA advertised over the sham link remains as Type 1 or Type 2. The MPLS VPN backbone or the backdoor link can be made the preferred path by adjusting the metrics. Figure 29 illustrates this scenario.

Figure 29 Using a Sham Link with a Backdoor Link



Using EIGRP for CE-PE Routing

EIGRP is another IGP used by many enterprise customers. This section specifically discusses what you might expect with EIGRP in an L3 MPLS environment, as well as backdoor connectivity for EIGRP. Backdoor connectivity between customer sites often causes problems for enterprises that run EIGRP as their PE-CE protocol. Typically, the intention is to use the L3 MPLS VPN service as the primary means of connectivity and the backdoor as a backup link. However, in backdoor scenarios with EIGRP as the PE-CE protocol and EIGRP over the backdoor link, the backdoor link may be preferred without some manipulation of the customer route. A feature known as BGP Cost Community can be implemented on the PE routers to add an additional comparison so that either choice is available based on configuration. EIGRP with and without BGP Cost Community support is discussed.

This section includes the following topics:

- What You Can Expect When Running EIGRP in a L3 MPLS VPN Environment, page 41
- EIGRP as PE-CE Backdoor Without Cost Community, page 44
- EIGRP PE-CE Backdoor with Cost Community, page 45

What You Can Expect When Running EIGRP in a L3 MPLS VPN Environment

When using EIGRP as the CE-PE routing protocol as well as in the enterprise sites, it is important for the enterprise customer to understand how EIGRP behaves and is treated in the L3 MPLS VPN environment. The following points describe the general operation, followed by some example scenarios.

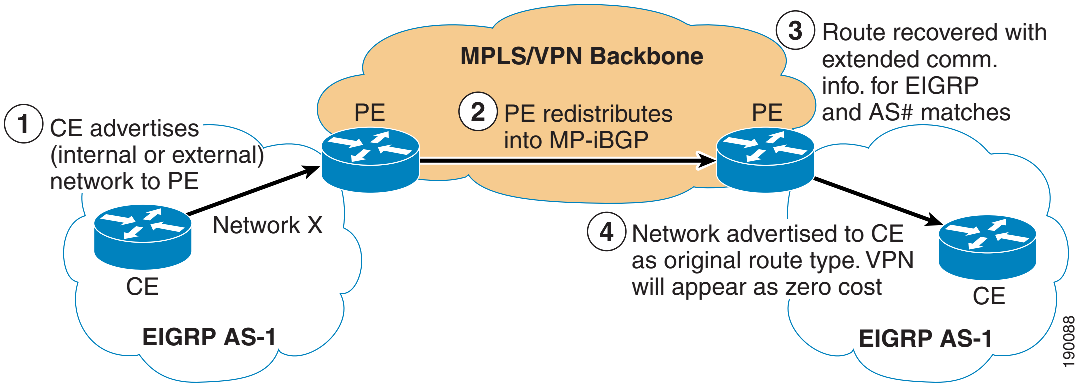
General operation is as follows:

- The CE router and the enterprise site run EIGRP as before. There is very little change needed, if any.
- The PE runs EIGRP in each VRF instance.
- The PE redistributes EIGRP into MP-iBGP.
- MP-iBGP carry extended community EIGRP information across the backbone to other sites.
- The PE routers at other sites redistribute the MP-iBGP routes back into EIGRP.

Example scenarios are as follows:

- Scenario 1—Enterprise sites are running EIGRP with the same EIGRP AS number. (See Figure 30.)

Figure 30 EIGRP—Same AS

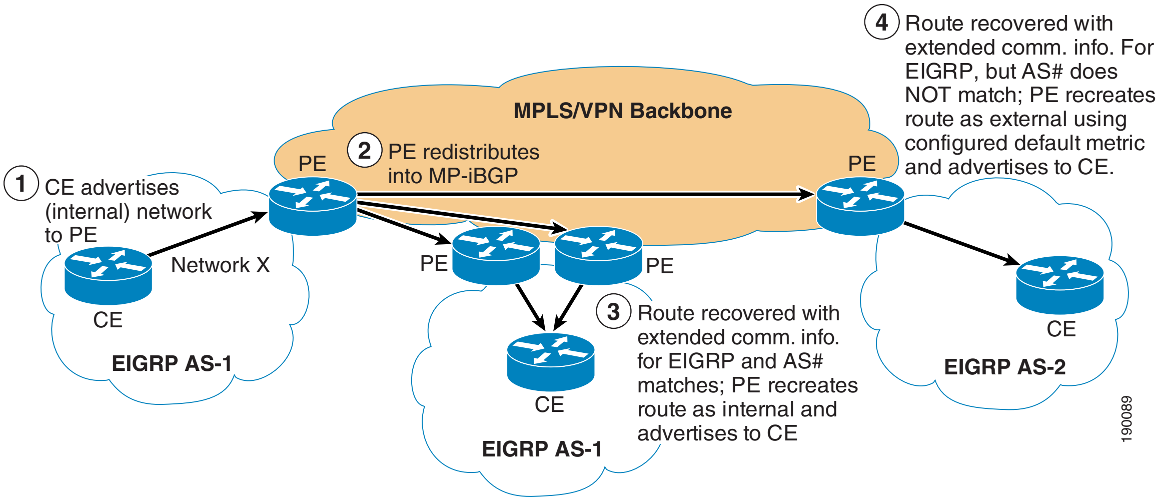

MPLS/VPN Backbone
1 CE advertises (internal or external) network to PE
PE
2 PE redistributes into MP-iBGP
PE
3 Route recovered with extended comm. info. for EIGRP and AS# matches
Network X
CE
4 Network advertised to CE as original route type. VPN will appear as zero cost
CE
EIGRP AS-1
EIGRP AS-1
190088

In Figure 30, the enterprise customer subscribed to the L3 MPLS VPN is running EIGRP at all their sites and using the same AS number at all their sites. The following sequence occurs:

1. The CE router advertises a network via EIGRP to the local PE router. This network can be internal or external.
2. The PE router redistributes the network from EIGRP into MP-iBGP with route information encoded in the extended community attributes within MP-iBGP.
3. The receiving PE receives the MP-iBGP updated that includes extended community information for EIGRP as well as a matching EIGRP AS number.
4. The PE recreates the EIGRP routes and sends them to the CE. These routes have the same route type as the original routes and they have the same metric as the sending PE had for these routes (the VPN backbone appears as zero cost).

- Scenario 2— Enterprise sites are running EIGRP, but not all sites use the same EIGRP AS number. (See Figure 31.)

Figure 31 EIGRP—Different AS

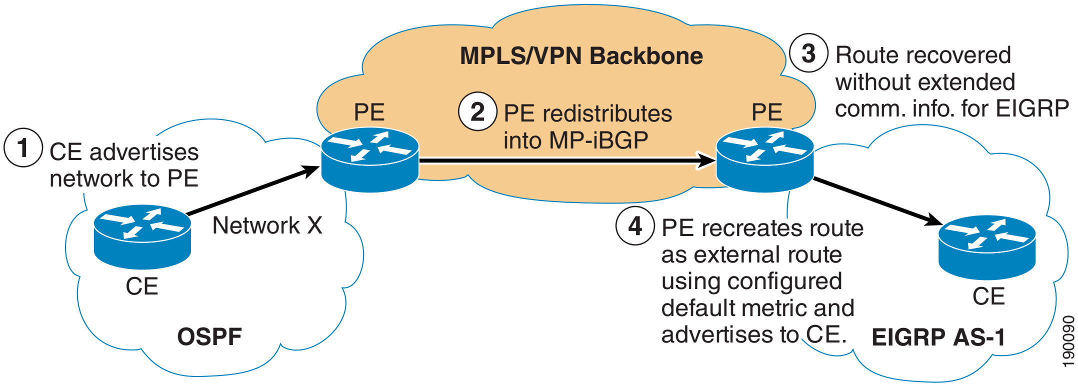

MPLS/VPN Backbone
1 CE advertises (internal) network to PE
2 PE redistributes into MP-iBGP
3 Route recovered with extended comm. info. for EIGRP and AS# matches; PE recreates route as internal and advertises to CE
4 Route recovered with extended comm. info. For EIGRP, but AS# does NOT match; PE recreates route as external using configured default metric and advertises to CE.
PE
PE
PE
PE
CE
CE
CE
Network X
EIGRP AS-1
EIGRP AS-1
EIGRP AS-2
190089

In Figure 31, the enterprise customer subscribed to the L3 MPLS VPN service is running EIGRP at all their sites, but not all sites are using the same EIGRP AS number. The following sequence occurs:

1. The CE router advertises a network via EIGRP to the local PE router. In this case, it can be said that the route is an internal route.
2. The PE router redistributes the network from EIGRP into MP-iBGP with route information encoded in the extended community attributes within MP-iBGP.
3. The PE(s) attached to EIGRP sites, using the same EIGRP AS number as the originating site, receives the MP-iBGP update, which includes extended community information for EIGRP as well as a matching EIGRP AS number. The receiving PE recreates the route as an internal route with the same metric as the sending PE had for this network.
4. The PE(s) attached to EIGRP sites, using a different EIGRP AS number as the originating site, receives the MP-iBGP update, which includes extended community information for EIGRP and a non-matching EIGRP AS number. The PE recreates the EIGRP route as an *external* EIGRP route using the configured default metric and advertises to the CE.

- Scenario 3—Some enterprise sites are running EIGRP while others are not. (See Figure 32.)

Figure 32 EIGRP—No EIGRP


MPLS/VPN Backbone
1 CE advertises network to PE
2 PE redistributes into MP-iBGP
3 Route recovered without extended comm. info. for EIGRP
4 PE recreates route as external route using configured default metric and advertises to CE.
PE
PE
CE
CE
Network X
OSPF
EIGRP AS-1
190090

In Figure 32, the enterprise customer subscribed to the L3 MPLS VPN service is running EIGRP at some sites but not at others. In this case, the customer is running OSPF at one site. The following sequence occurs:

1. The CE router advertises a network via OSPF to the local PE router.
2. The PE router redistributes the network from OSPF into MP-iBGP with route information encoded in the extended community attributes within MP-iBGP.
3. The PE attached to EIGRP site receives the MP-iBGP update, which does *not* include any extended community information for EIGRP.
4. The receiving PE, attached to the EIGRP site, recreates the route as an *external* EIGRP route using the configured default metric and advertises it to the CE. The originating protocol appears to be BGP.

As can be seen by the above scenarios, EIGRP behaves slightly differently based on the scenario.

Auto-Summarization

When configuring EIGRP, the default EIGRP behavior is to enable auto-summarization. Auto-summarization causes EIGRP to automatically summarize a network when a major network boundary is crossed. This might have undesirable effects on the network in an MPLS VPN environment. A site could receive the same summary from multiple other sites and would not be able to determine which site to use for a more specific route. If proper site addressing is planned, then auto-summarization should be disabled and any summarization can be manually performed.

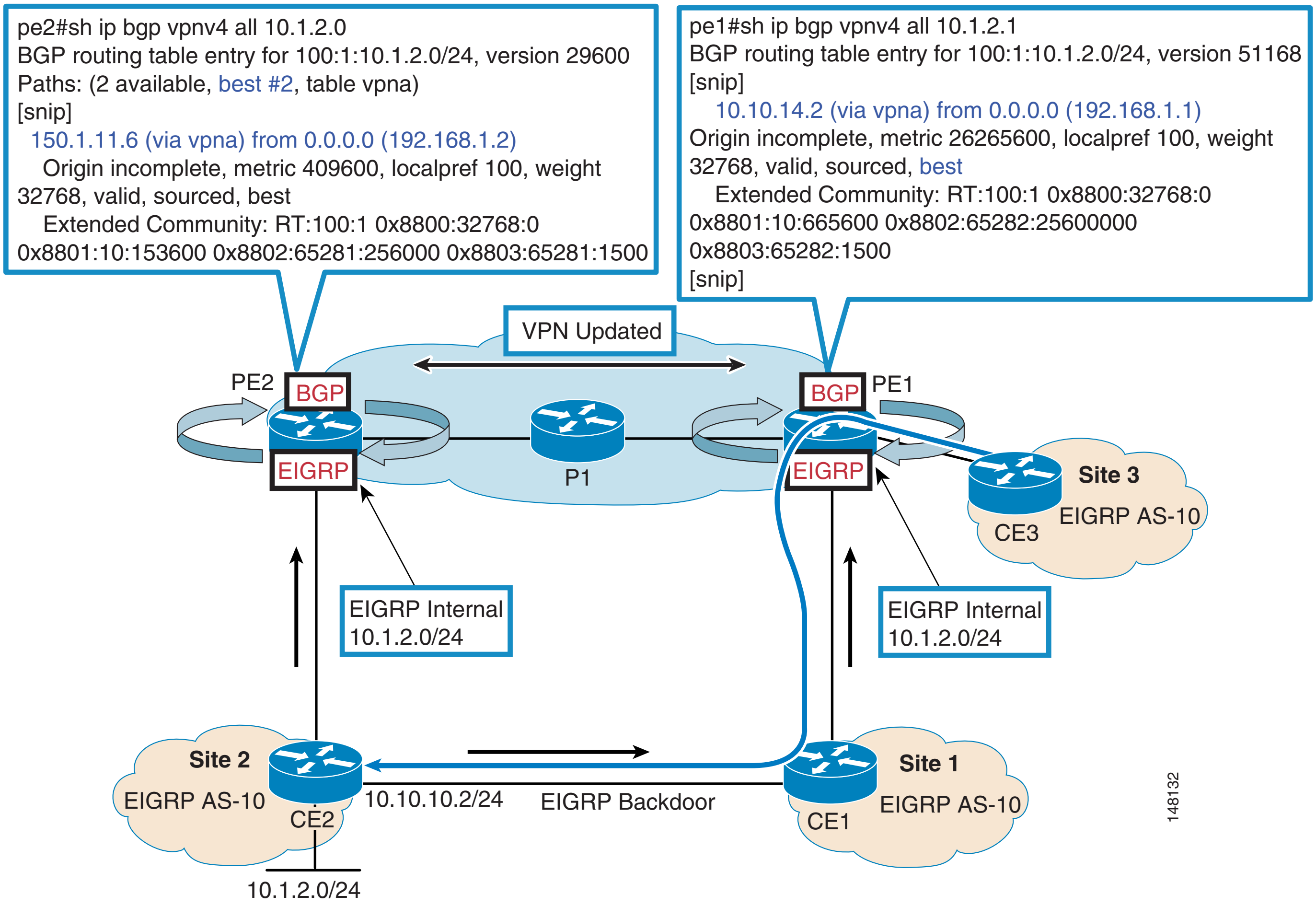
EIGRP as PE-CE Backdoor Without Cost Community

Figure 33 illustrates a typical backdoor scenario with EIGRP as the PE-CE protocol and IGP over the backdoor link. Prefix 10.1.2.0/24 is owned by Site2 and advertised in EIGRP. The EIGRP route is advertised to CE1 in Site1 and PE2. PE2 installs this route in its VRF, redistributes this EIGRP route into BGP, and passes the route to PE1. Likewise, CE1 advertises this route to PE1. PE1 has two BGP paths available for 10.1.2.0/24:

- The iBGP advertisement from PE2
- The locally redistributed BGP route from the CE1 EIGRP advertisement.

In this case, a locally originated route is preferred based on the BGP best path decision process. However, such a decision leads to traffic being forwarded over the backdoor link as its primary path. For example, traffic originated in Site3 is destined to 10.1.2.0/24 routes to Site1 and over the backdoor link to Site2, as illustrated by the red arrow in Figure 33.

Figure 33 EIGRP PE-CE Backdoor Scenario without BGP Cost Community



```
pe2#sh ip bgp vpnv4 all 10.1.2.0
BGP routing table entry for 100:1:10.1.2.0/24, version 29600
Paths: (2 available, best #2, table vpna)
[snip]
 150.1.11.6 (via vpna) from 0.0.0.0 (192.168.1.2)
   Origin incomplete, metric 409600, localpref 100, weight
32768, valid, sourced, best
   Extended Community: RT:100:1 0x8800:32768:0
0x8801:10:153600 0x8802:65281:256000 0x8803:65281:1500
```

```
pe1#sh ip bgp vpnv4 all 10.1.2.1
BGP routing table entry for 100:1:10.1.2.0/24, version 51168
[snip]
  10.10.14.2 (via vpna) from 0.0.0.0 (192.168.1.1)
Origin incomplete, metric 26265600, localpref 100, weight
32768, valid, sourced, best
  Extended Community: RT:100:1 0x8800:32768:0
0x8801:10:665600 0x8802:65282:25600000
0x8803:65282:1500
[snip]
```

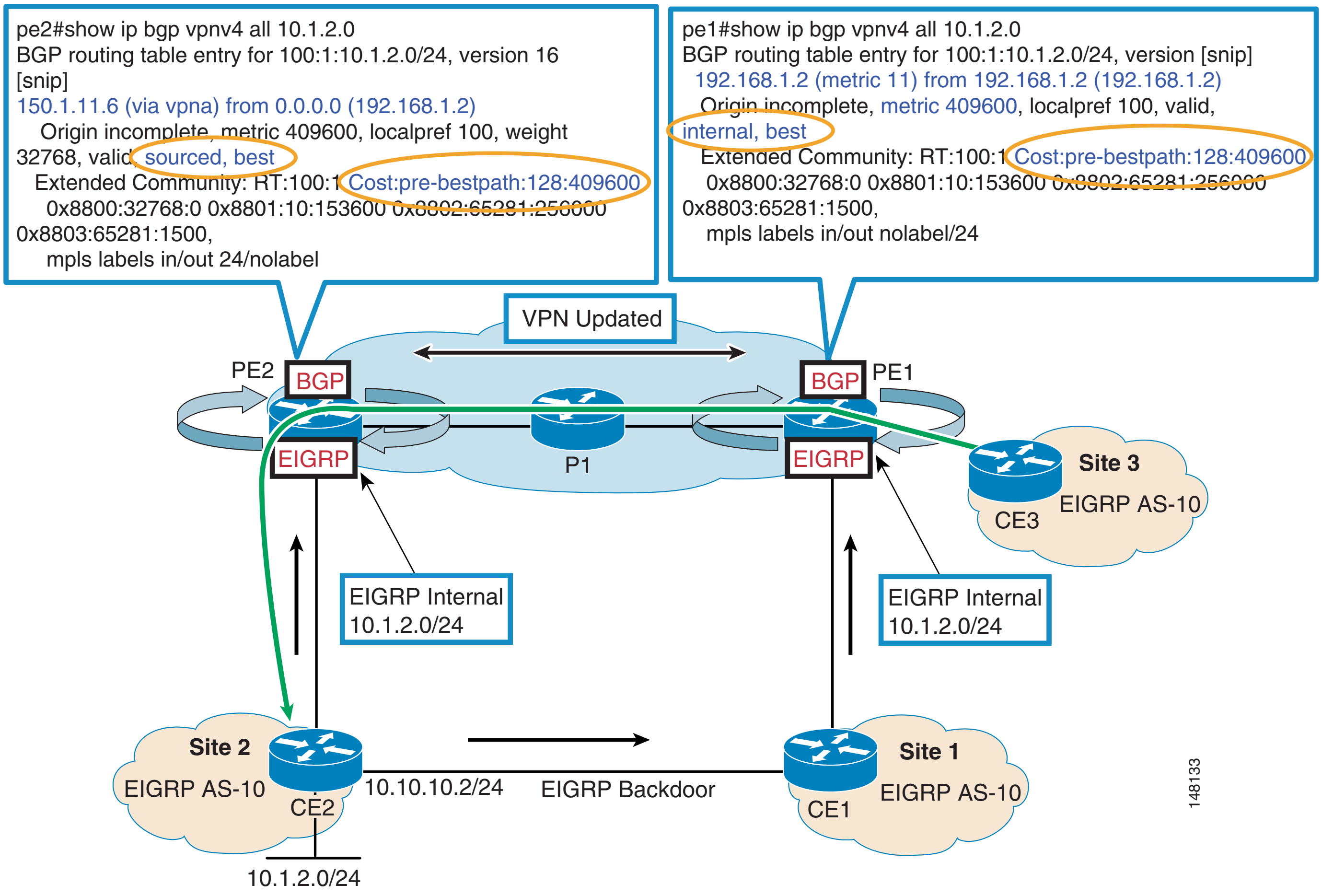
EIGRP PE-CE Backdoor with Cost Community

The topology shown in Figure 33 does not achieve the goal of using the MPLS VPN service as a primary path. A means of affecting the PE1 choice must be available to have it prefer the iBGP learned announcement from PE2. This can be achieved through manipulation of BGP attributes or metrics before the local originated tiebreaker, such as weight or local preference. However, an additional attribute known as BGP Cost Community was developed to handle this case in a more graceful manner. More information regarding this feature can be found at the following URL: http://www.cisco.com/en/US/docs/ios/12_2s/feature/guide/fsbgpcce.html.

The BGP Cost Community feature is configured on the PE and attaches an extended community attribute. This cost community value is compared, and it influences the path determination. By adjusting these community values as needed, traffic can be forwarded to the correct path.

Figure 34 illustrates how BGP Cost Community can cause the PE routers to prefer the L3 MPLS VPN service as the primary path. By default, when PE2 redistributes the EIGRP path into BGP, the BGP Cost Community attribute is populated with the EIGRP metric. As in Figure 33, PE-1 has two options: the iBGP learned path from PE-2 or the locally originated BGP path learned through redistribution of the EIGRP route from CE-1. Because the EIGRP metric of the route advertised from CE-1 includes the added cost of traversing the backdoor link, the BGP Cost community of the iBGP path is lower, and thus preferred and installed. As such, the traffic from Site3 destined to 10.1.2.0/24 is forwarded over the MPLS VPN service instead of the backdoor link. This is illustrated by the green arrow in Figure 34.

Figure 34 EIGRP PE-CE Backdoor Scenario with BGP Cost Community



```
pe2#show ip bgp vpnv4 all 10.1.2.0
BGP routing table entry for 100:1:10.1.2.0/24, version 16
[snip]
150.1.11.6 (via vpna) from 0.0.0.0 (192.168.1.2)
   Origin incomplete, metric 409600, localpref 100, weight
32768, valid, sourced, best
  Extended Community: RT:100:1 Cost:pre-bestpath:128:409600
    0x8800:32768:0 0x8801:10:153600 0x8802:65281:256000
0x8803:65281:1500,
    mpls labels in/out 24/nolabel
```

```
pe1#show ip bgp vpnv4 all 10.1.2.0
BGP routing table entry for 100:1:10.1.2.0/24, version [snip]
  192.168.1.2 (metric 11) from 192.168.1.2 (192.168.1.2)
   Origin incomplete, metric 409600, localpref 100, valid,
internal, best
   Extended Community: RT:100:1 Cost:pre-bestpath:128:409600
    0x8800:32768:0 0x8801:10:153600 0x8802:65281:256000
0x8803:65281:1500,
    mpls labels in/out nolabel/24
```

Default Route Handling

This section describes other routing considerations that are important when implementing an L3 MPLS VPN. It includes the following topics:

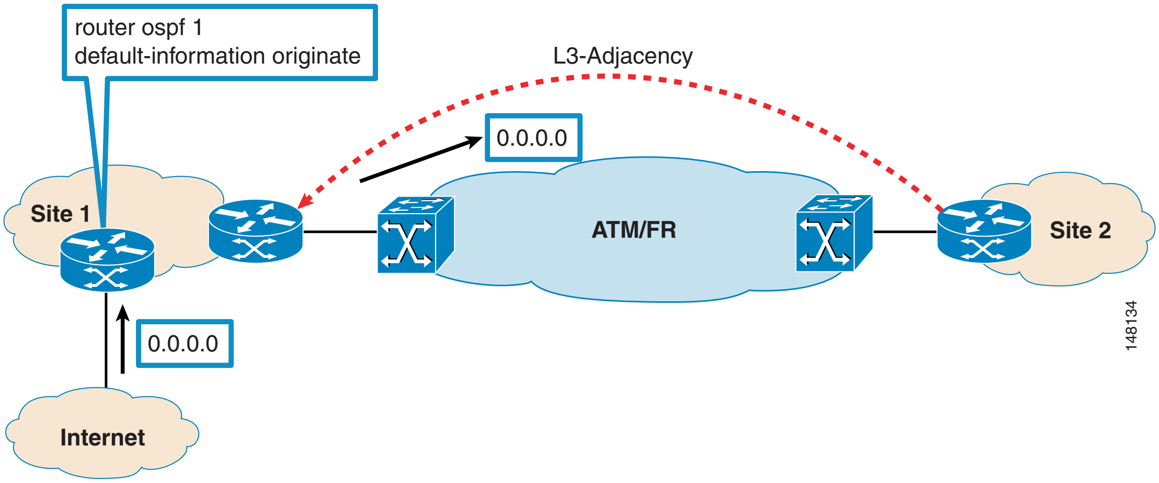
- Default Route Handling Overview, page 47
- Default Routing in a Multihub Environment, page 49
- Handling Multiple Default Routes with IGP as PE-CE Protocol, page 51
- Handling Multiple Default Routes with BGP as PE-CE Protocol, page 52

Default Route Handling Overview

In the MPLS VPN environment, you need to be careful about propagating the default route from one site to other site across the MPLS VPN cloud, especially if the PE-CE protocol is an IGP such as OSPF or EIGRP. For example, Site1 is a hub or an Internet gateway site and learns a default route from an ISP, either dynamically or statically. To inject the default route locally in the OSPF domain, it is normally explicitly injected in the site using the **default-information originate** command in OSPF.

Figure 35 below shows two enterprise network sites connected with each other over traditional Layer 2 ATM/Frame Relay links. The default route is originated at the gateway router at the hub location (Site1), which is then flooded everywhere in the OSPF domain, including Site2. The underlying ATM/Frame Relay cloud does not participate in the routing, and the enterprise has full control of distributing the default route information to the network.

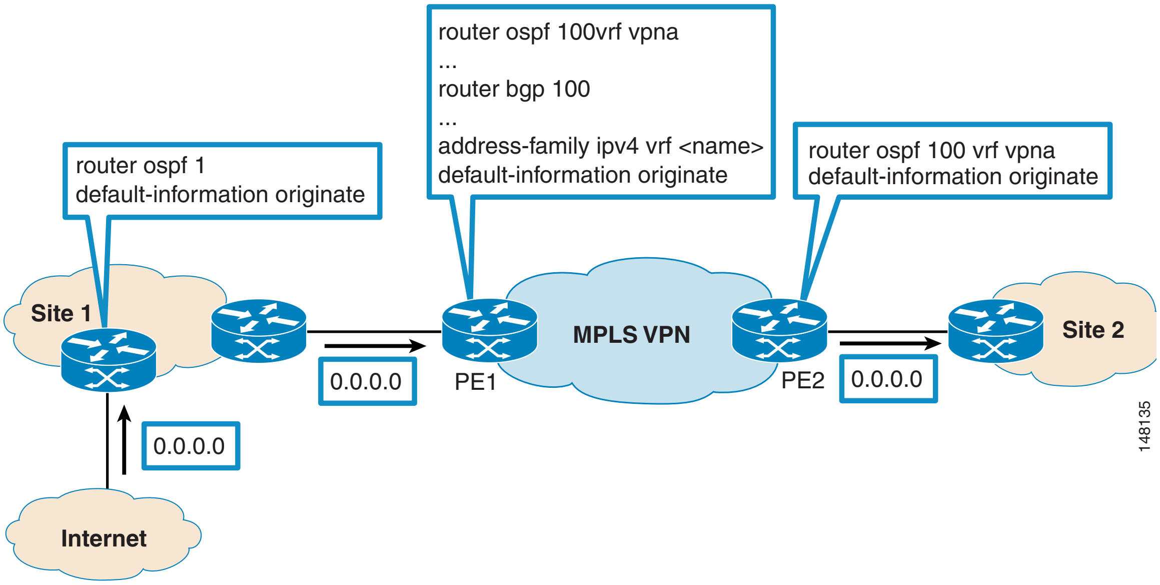
Figure 35 Originating the Default Route in the Traditional L2 WAN Core



However, in the MPLS VPN environment, default route propagation from one site to the other is not transparent. A default route originated at the gateway router in Site1 is not automatically propagated to Site2. This is because PE routers in the MPLS VPN cloud exchange CE routing information using MP-BGP, and a default route from a non-BGP PECE protocol such as OSPF is not injected into BGP via redistribution. Instead, it needs to be explicitly re-originated in the MPLS VPN cloud in BGP within the "vpna" context on PE1 and within the "vpna vrf OSPF" context on PE2. This is illustrated in Figure 36.

It is therefore important that you inform the SP of the sites that originate the default route so that the necessary configuration is put in place to distribute the default route appropriately to the other sites.

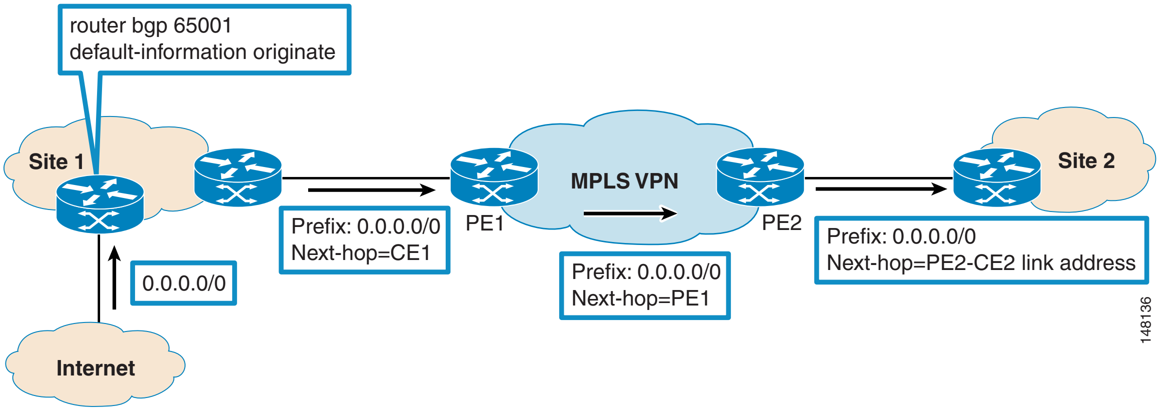
Figure 36 Originating Default in the MPLS VPN Environment



When BGP is used as the PE-CE protocol, no additional configuration on PE1 or PE2 is needed to propagate the default to other PEs within the MPLS cloud. A BGP-learned default is automatically advertised to all the other BGP neighbors unless explicitly filtered.

Figure 37 shows the default route propagation from Site1 to Site2. Note that the **default-information originate** command is needed on the Internet gateway router only if the default is learned from a protocol other than BGP. In this example, a static default is configured and then originated within BGP using the **default-information originate** command.

Figure 37 Default Route Propagation using BGP as PE-CE



Default Routing in a Multihub Environment

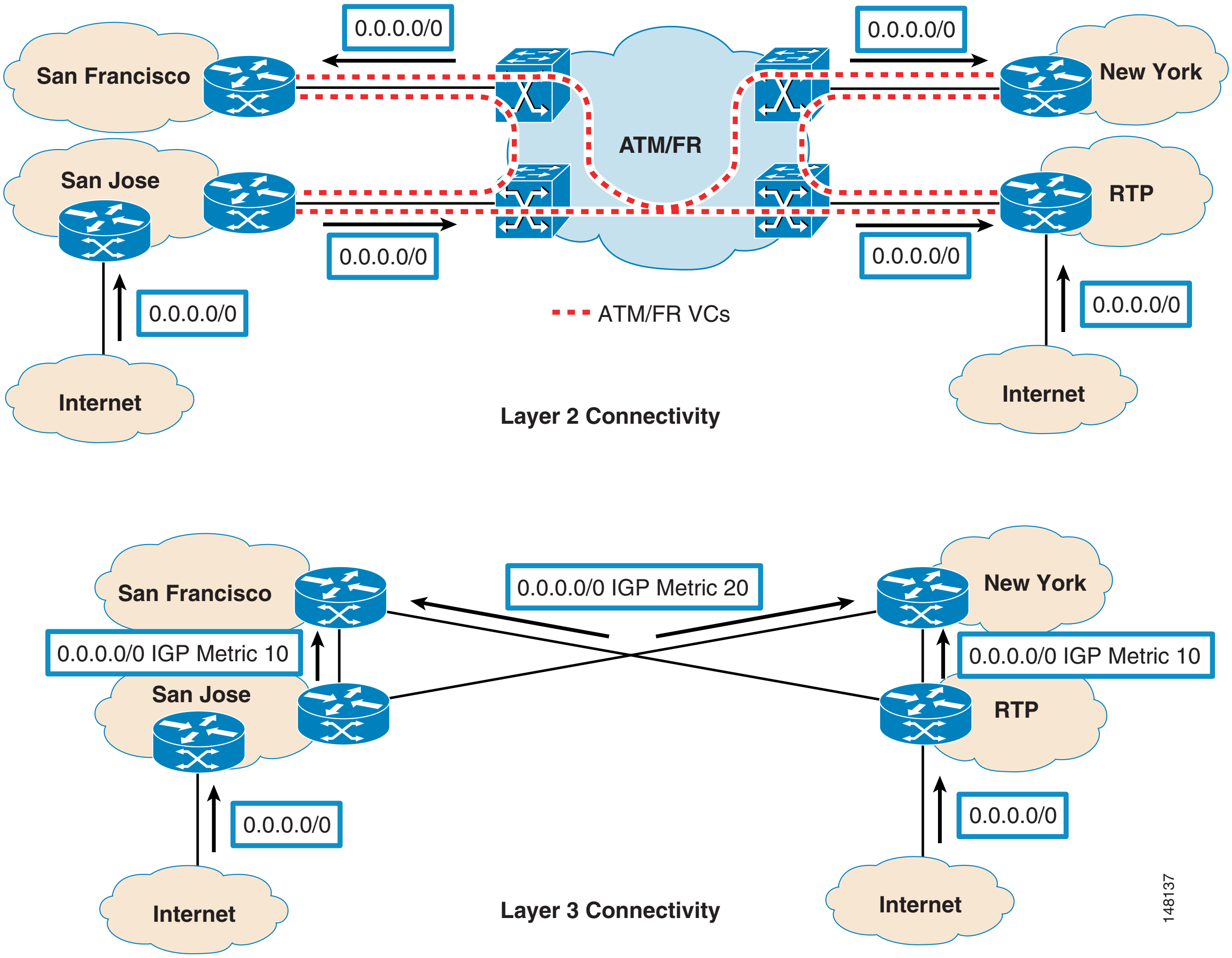
In enterprise networks, multiple Internet gateways or mirrored data center locations that inject a default route in the network are often used for redundancy. To optimize network resource use, the traffic load is often distributed across multiple exit points in the network. In addition, it is desirable for each gateway to back up the other and for the traffic to be rerouted to the alternate gateway if connectivity problems occur.

In the traditional enterprise networks, the customer IGP usually determines the closest data center (hub site) or the closest Internet gateway location. As illustrated in Figure 38, San Jose and RTP are two hub or Internet gateway sites that receive a default route.

Based on the IGP metric, San Francisco uses San Jose to go to the Internet, while New York prefers the RTP exit. In case San Jose loses Internet connectivity, San Francisco determines that RTP is the best path to exit to the Internet.

In this environment, the enterprise customer can design optimal paths for Internet traffic. By manipulating the IGP metric, either San Jose or RTP can be made the preferred exit point for Internet traffic. If stateful firewalls are used, BGP can enforce symmetric routing.

Figure 38 Multiple Defaults in a Layer 2-Based VPN

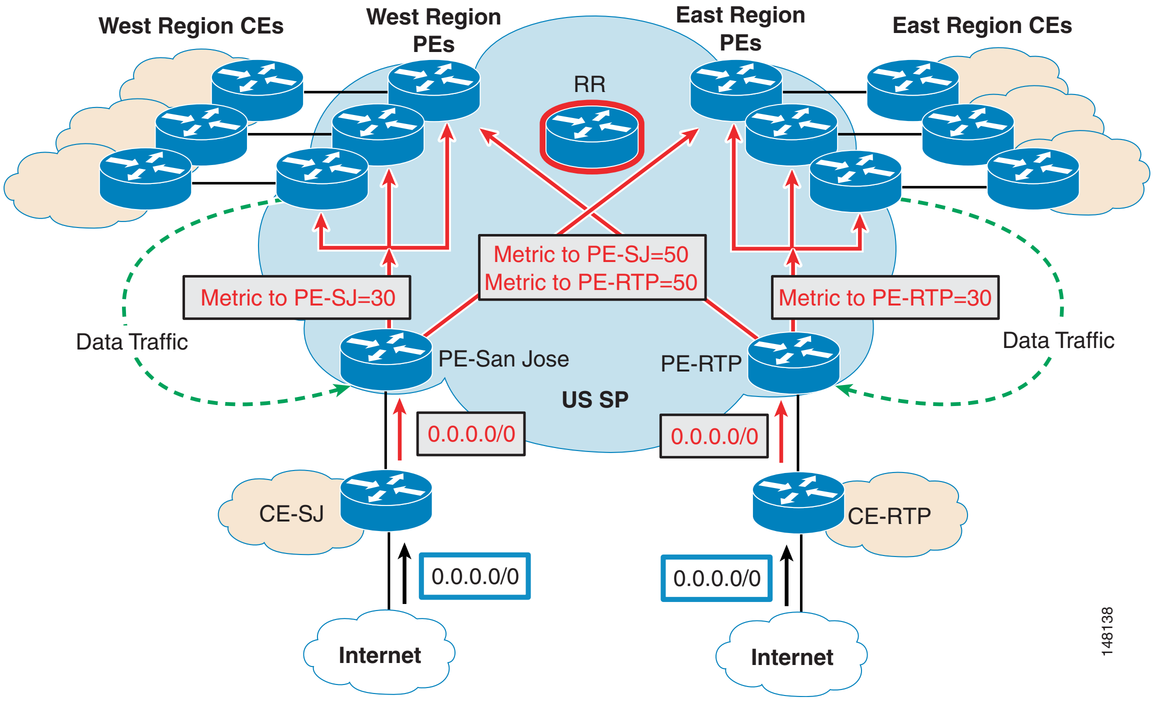


However, when the existing WAN core is migrated to MPLS VPN service, it is not the customer IGP that determines the optimal routing across the MPLS cloud. Instead, this usually depends on the routing policies inside the SP network.

Figure 39 illustrates an example where a US enterprise that has migrated to MPLS VPN service for WAN connectivity requires the following routing policies to be enforced across various sites:

- Two sites (San Jose and RTP) advertise 0.0.0.0/0 routes to the spoke sites
- For Internet access, satellite sites in the West Region should follow the default route learned from San Jose; East Region sites should use RTP.
- In case of failure, spoke sites should use the non-preferred default route. In other words, the West Region may follow the default from RTP if San Jose loses its connectivity to the Internet. The East Region may exit from San Jose if RTP loses its default route to the Internet.

Figure 39 Default Behavior With Multiple Defaults in an MPLS VPN Environment



Assuming that no route reflectors are used in the SP network, all the PEs receive default routes from both RTP and SJC locations. Within the SP cloud, the best path decision is based on the BGP attributes. In this example, it is determined by lowest IGP metric to the next hop. As a result, it is possible that West Region and East Region PEs may use San Jose and RTP sites respectively to get to the Internet as required.

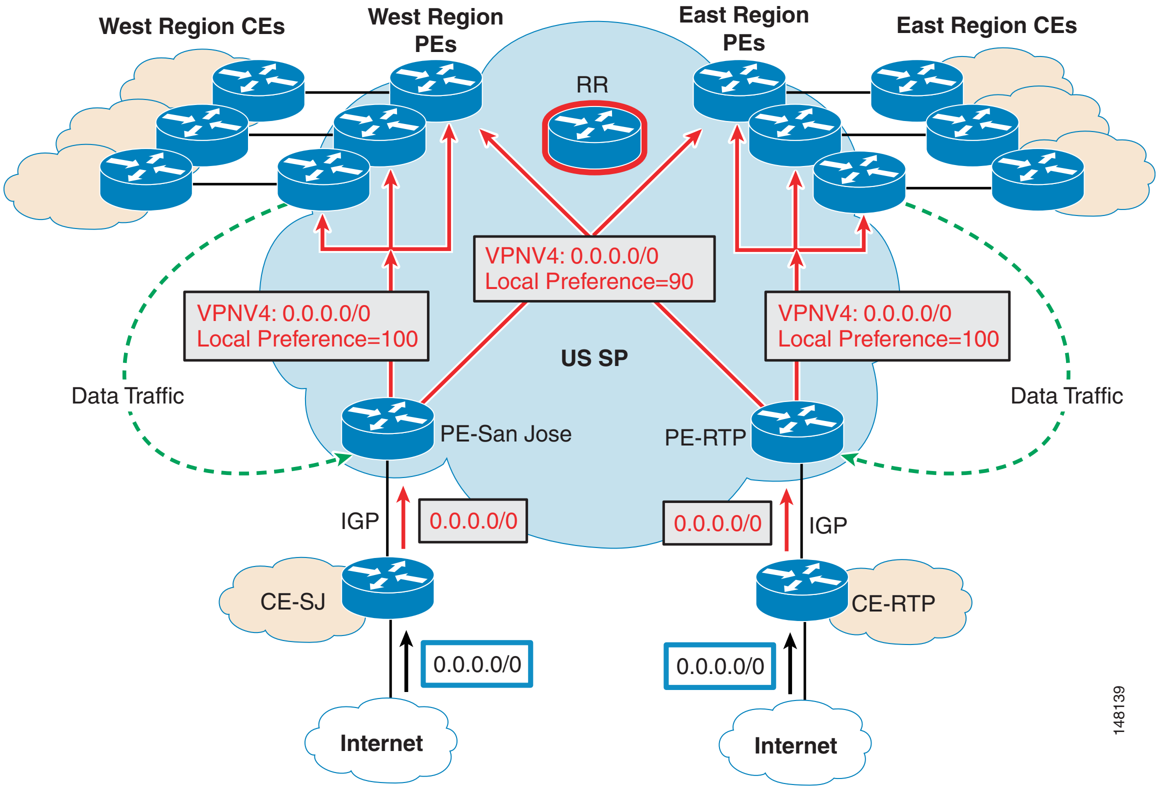
Note that in case of a link or node failure, IGP metric changes within the SP cloud can potentially have an impact on the customer traffic destined to the Internet. In addition, this example assumes that all the customer sites are connected to either the West Region or East Region PEs and appear to be following the optimal path in the steady state conditions. However, in the presence of other customer locations that are connected to the SP cloud elsewhere, traffic may not use the desired path. For example, if customer traffic in Kansas is supposed to follow the San Jose exit but the SP PE in Kansas sees a better route metric towards the East Region, RTP is selected as the best path. In other words, a path that is considered optimal for the SP is not necessarily an optimal path for the enterprise customer, and you need to keep this in mind.

Handling Multiple Default Routes with IGP as PE-CE Protocol

If the PE-CE protocol is IGP, there is not much an enterprise can do to dynamically influence the default route preference within the MPLS VPN SP core. The only solution is for the SP to implement BGP policies, such as setting the appropriate local preference, to achieve the desired routing behavior before the routes are propagated to the other PEs within the cloud. However, this requires SP configuration for each customer and PE neighbor.

Figure 40 shows an example with local preference.

Figure 40 Influencing Default Route Preference by Adjusting BGP Attributes



To achieve the desired routing behavior, the default route must be advertised in such a way that East Region PEs receive a lower local preference value (90 in this example) than San Jose, and higher local preference (default=100 in this example) than RTP. Similarly, West Region PEs must receive a default route with a lower local preference (90) value than RTP, and higher than SJ (100=default). In this scenario, if San Jose loses its default route, West Region PEs can revert to the RTP route.

In the examples so far, it is assumed that route reflectors do not exist in the SP cloud. However, route reflectors are almost always present in large networks, as shown in Figure 40. Like any BGP speaking router, if a route reflector receives more than one route for the same prefix, it selects one best path using the BGP best path algorithm.

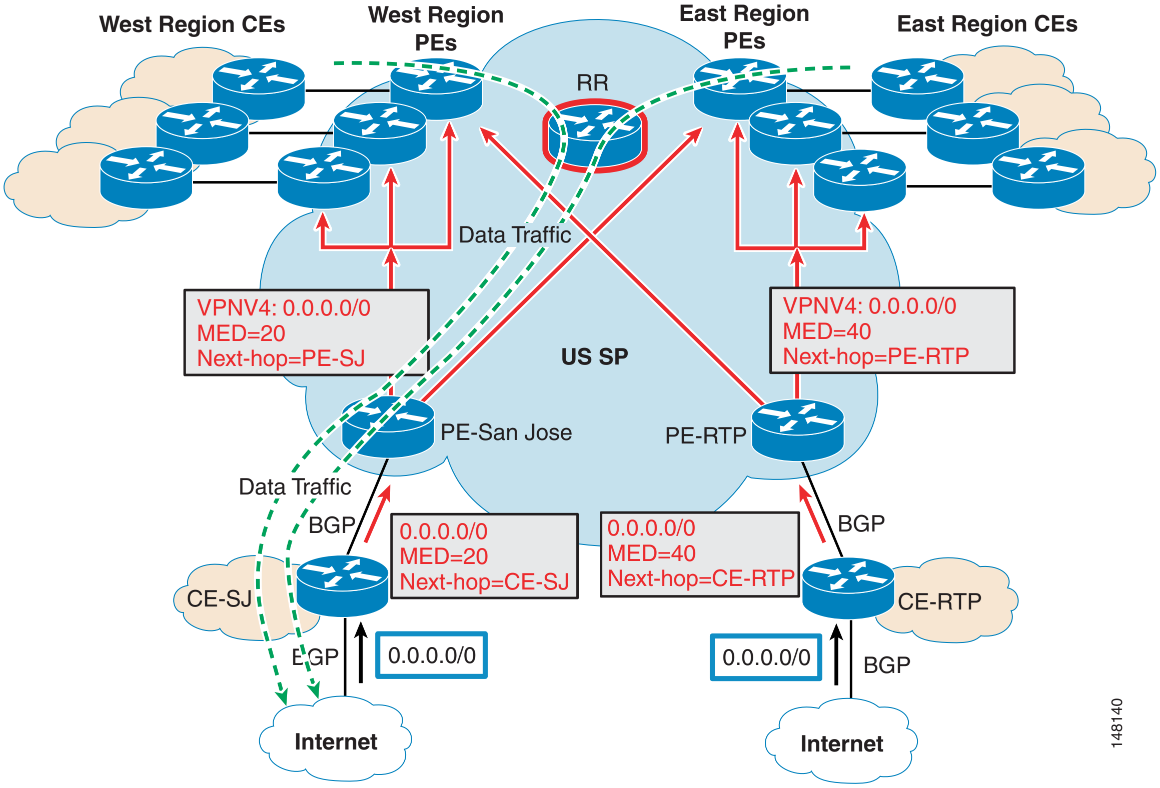
For example, assume a route reflector selects the default route from San Jose, because of better local preference, and reflects it to all the PEs including the East Region and the West Region PEs. As a result, traffic from all the sites destined to the Internet exits through San Jose instead of being distributed between the two gateway locations. When the default route from San Jose is lost, all the traffic then switches over to follow the default learned through RTP. In addition, in this scenario, it is not possible to apply per-neighbor policies

To achieve the desired routing behavior in the presence of route reflectors, the SP must use different RD values on each PE for the customer VPN. This makes the route reflector reflect both default routes (SJ and RTP) to all the PEs. The PEs can make their own best path selection based on local preference, IGP metric, or other attributes associated with the routes.

Handling Multiple Default Routes with BGP as PE-CE Protocol

CE routers can influence the routing behavior across the MPLS VPN cloud to some extent. For example, it is not feasible for the customer routers to dictate that East Region PEs should use RTP and West Region PEs should use San Jose. However, it is possible to dynamically indicate to the SP cloud that San Jose is the primary and RTP is the secondary gateway site (see Figure 41).

Figure 41 Influencing the Default Route Preference with Customer MED Values



All Internet traffic can transit through San Jose in the steady state conditions, and RTP can be used to access Internet destinations if San Jose fails. One way to achieve this effect is to do the following:

- Advertise the default route from SJ-CE to SJ-PE with a lower MED value.
- Advertise the default route from RTP-CE to RTP-PE with a higher MED value.
- Use different RDs for the customer VRF on San Jose and RTP PEs.
- If San Jose and RTP are using different AS numbers, enter the **bgp compare-med always** command on the PEs so that MEDs from different ASes can be compared.

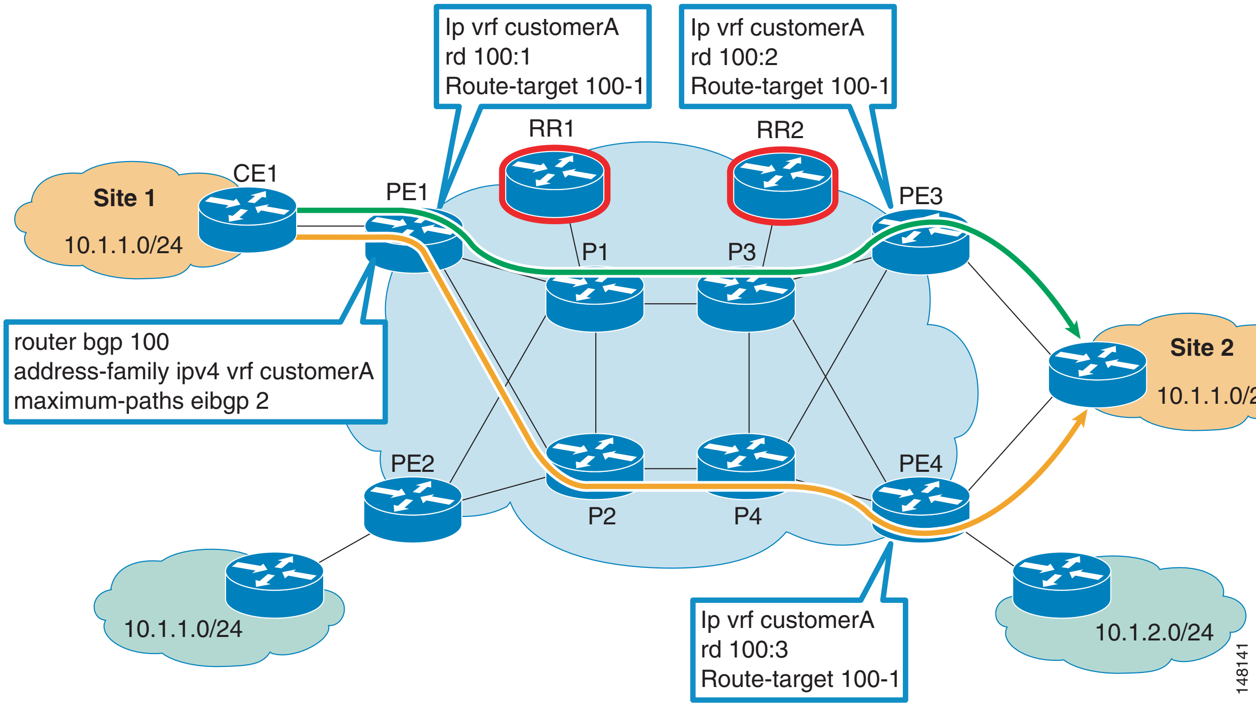
The MED attribute is used to influence the return traffic but other attributes such as AS path prepend can also be used to achieve the same effect.

Load Balancing

Multihomed CEs often require load balancing across multiple available paths. As mentioned earlier, in a full mesh BGP environment, PEs receive all the available paths to a given prefix and load balancing can easily be achieved. However, when route reflectors are present in the SP core, PE routers receive only one route even if multiple paths exist, and load balancing does not occur.

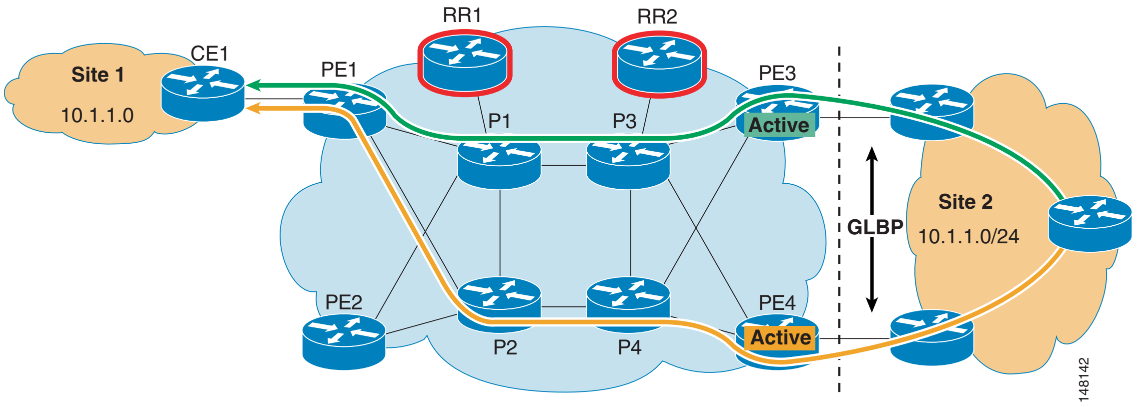
To achieve load balancing, the SP needs to implement unique RD values for the customer VPN on each PE router. In addition, eiBGP configuration with the desired number of paths (across which load balancing is desired) needs to be enabled in the SP environment. Figure 42 illustrates a load balancing example.

Figure 42 Load Balancing using Different RDs



If the PE-CE link is Ethernet-based, HSRP/VRRP or GLBP can be used for redundancy. By default, HSRP or VRRP do not load balance across the two edge routers. GLBP provides an effective alternative solution that can be used to achieve true load balancing, as shown in Figure 43.

Figure 43 Load Balancing with GLBP



Multihoming Scenarios

Requirements for multihoming may vary, depending on the goals of the IP-VPN customer.

One of the main goals of multihoming is to provide availability by means of redundancy, which can mean different things to different people. This redundancy can occur at the port level, line card level, box level, site level, and so on. Each scenario comes with both financial cost and technical complexity; however, they do not reduce the requirement for multihoming for IP-VPN customers. This section describes the technical challenges caused by various multihoming scenarios.

The most important multihoming question is whether to choose a single provider or multiple VPN providers. The technical answer to this question depends of the capabilities of the VPN provider, such as (but not limited to) the following:

- Geographical reach
- Ability to offer consistent routing and QoS services that meet the customers requirements
- Security best practices,
- Technical support and monitoring of the infrastructure

When these questions are answered, and after doing a financial comparison between the two options, VPN customers can choose either single or dual VPN providers for multihoming their VPN sites.

The following sections discuss the most common technical challenges that VPN customers face with single or dual IP-VPN provider multihoming scenarios.

Single Provider

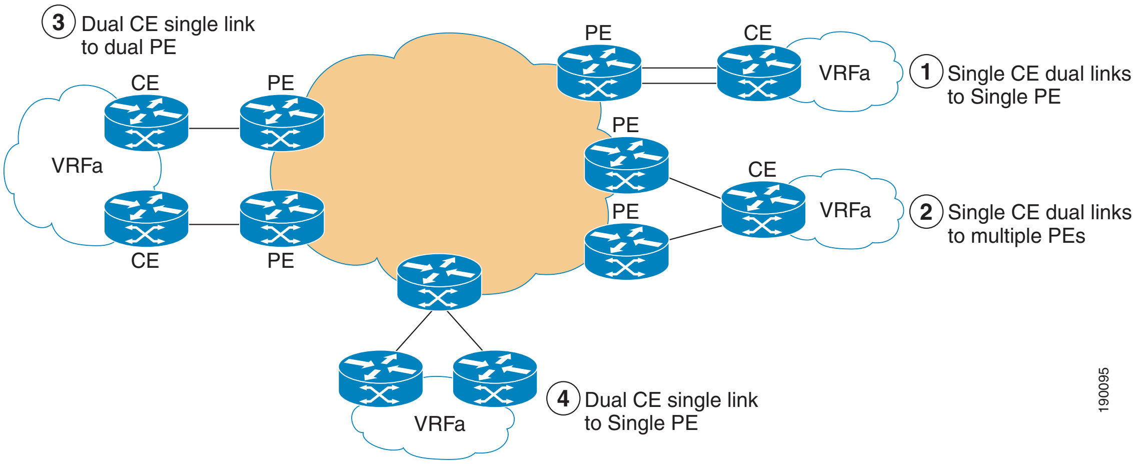
The following are five single provider scenarios:

1. Single CE with dual links to a single PE
2. Single CE with dual links to multiple PEs
3. Dual CE with a single link to a dual PE
4. Dual CE with a single link to a single PE
5. Backup using DMVPN over the Internet

Scenarios 1–4

Figure 44 shows the most common topologies when multihoming with single provider VPN service.

Figure 44 Multihoming—Single Provider


3 Dual CE single link to dual PE
CE
PE
VRFa
CE
PE
PE
CE
VRFa
1 Single CE dual links to Single PE
PE
CE
PE
VRFa
2 Single CE dual links to multiple PEs
VRFa
4 Dual CE single link to Single PE
190095

The enterprise customer must decide whether to load balance traffic across multiple PE-CE links, or to use them in primary/backup fashion. BGP is a very common PE-CE routing protocol. With BGP, features and attributes such as maximum path selection, local preference, and AS path length (among many others) can be manipulated to load balance traffic.

Note For detailed examples on load balancing with BGP in multihoming environments, see “Load Sharing with BGP in Single and Multihomed Environments” at the following URL: http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a00800945bf.shtml.

In the case where the PE-CE routing protocol is an IGP, the IGP metrics can be manipulated to achieve load balancing. Each IGP calculates its metric differently, so how the metric is manipulated depends on the routing protocol being used.

Note For more information on load balancing, see “How Does Load Balancing Work?” at the following URL: http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094820.shtml.

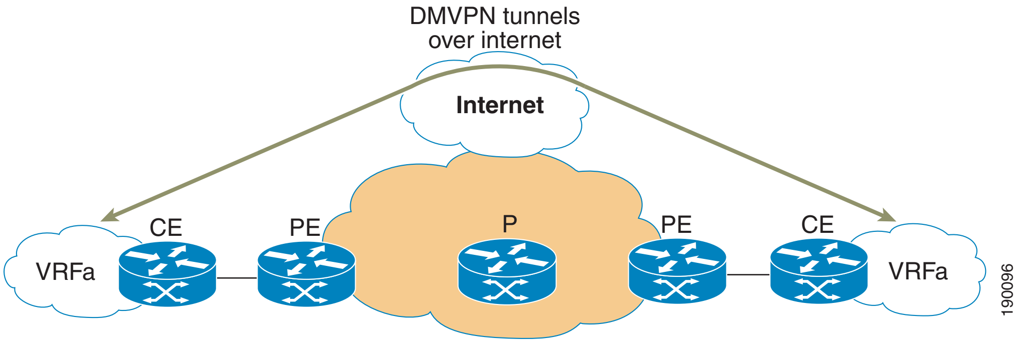
Scenario 5—Backup using DMVPN over the Internet

In this case, the customer buys only one VPN service from the provider and uses the Internet as a backup. It is possible to deploy DMVPN across the Internet to provide backup services in case the primary MPLS-VPN service becomes unavailable.

DMVPN has been in use for awhile, and enterprises have used it successfully to connect sites together over the Internet while securing the traffic using IPsec.

Figure 45 shows the network topology when both IP-VPN and DMVPN are used.

Figure 45 Scenario 5—Using DMVPN



Note Configuration of DMVPN can be found at the following URL: http://www.cisco.com/en/US/products/sw/iosswrel/ps1839/products_feature_guide09186a0080110ba1.html.

Typically, an IGP such as OSPF or EIGRP runs on top of DMVPN to propagate enterprise IP networks. Among other considerations, the following should be carefully examined:

- Explicitly configure DMVPN as a backup for the IP-VPN path.
- Avoid redistribution between PE-CE and DMVPN learned routes to protect against loops.

The following two scenarios should be considered in terms of routing preference:

- When the PE-CE routing protocol is identical to IGP running on DMVPN (EIGRP, OSPF, or RIP)

  To avoid making DMPVN a preferred path, make sure that IGP metrics are higher on the DMVPN tunnel interface as compared to PE-CE interface. For example, in the case of OSPF, using the **ip ospf cost <cost>** command on the tunnel interface can change the cost to reach remote sites.

- When the PE-CE routing protocol is EBGP and an IGP (EIGRP, OSPF, or RIP) running on the DMVPN.

  In Cisco IOS, administrative distance (AD) is the first tie breaker when more than one path exists for the same network and lowest is preferred. EBGP (20), EIGRP internal (90), EIGRP external (170), OSPF (110), and RIP (120) are the ADs used in Cisco IOS.

  With EBGP as the PE-CE protocol, it is always chosen as a preferred path over the IGP path learned over the DMVPN tunnel.

Dual Provider

This section describes the following two dual provider scenarios,:

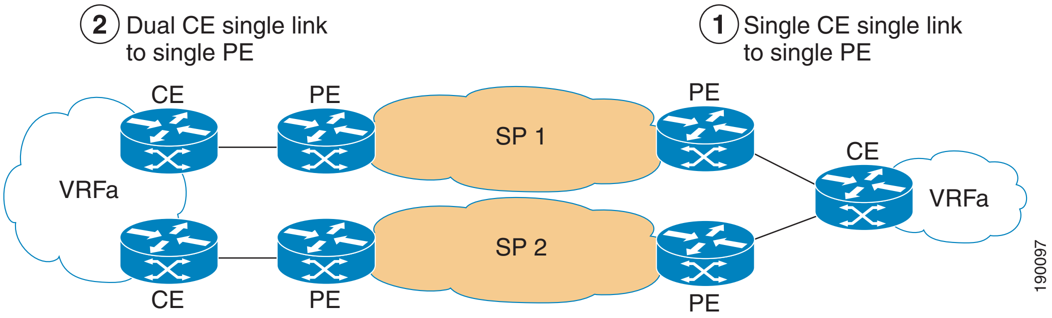
1. Single CE with a single link to a single PE
2. Dual CE with a single link to a single PE

These scenarios are followed by a sub-optimal routing scenario.

Note For more information, see the following URL: http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a008009456d.shtml

Figure 46 shows sample topologies illustrating scenarios 1 and 2 of dual provider VPN service.

Figure 46 Multihoming—Dual Provider


2 Dual CE single link to single PE
1 Single CE single link to single PE
CE
PE
SP 1
PE
CE
VRFa
VRFa
SP 2
CE
PE
PE
190097

As previously mentioned, it is the responsibility of the enterprise to decide whether to load balance traffic across multiple PE-CE links, or to use them in primary/backup fashion. Primary/backup routes can be achieved using similar techniques as those discussed in the single provider multihomed topologies section above. However, load balancing across two providers can be complicated. If EBGP is the choice of PE-CE from both providers, BGP updates from both providers are different because of AS_PATH, and are thus subject to a single best path calculation. In such a case, this enterprise must resolve to a standard prefix splitting preference as described in the document at the following URL: http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a00800945bf.shtml.

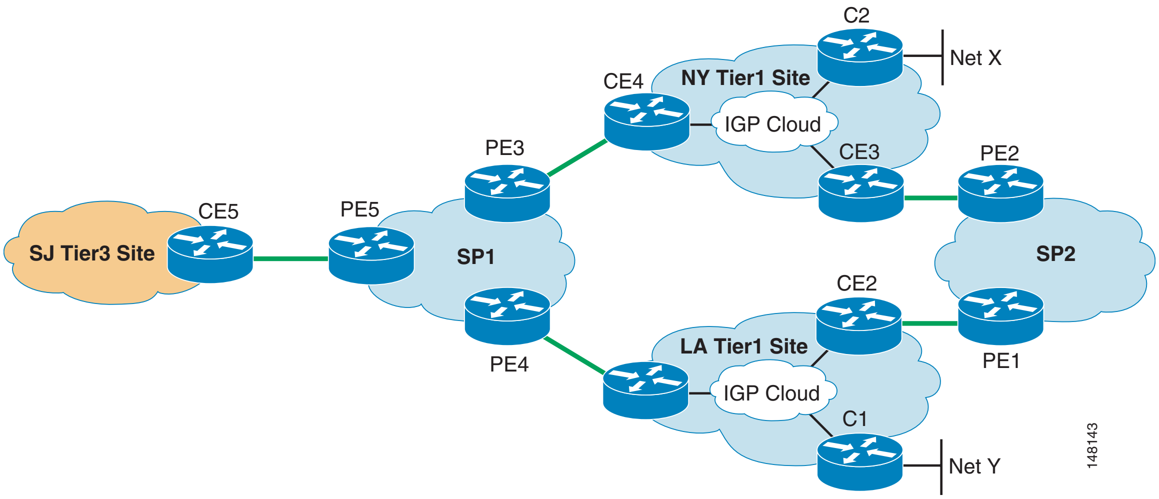
If a common IGP (OSPF, EIGRP, or RIP) is used as the PE-CE routing protocol for both providers, load sharing takes place by default provided both paths have equal cost.

Note Choose CE routers appropriately to handle multiple routes, configuration, and so on, when multihomed.

Sub-optimal Routing Scenario

For redundancy purposes, an enterprise might choose multiple providers for their L3 VPN services. The enterprise may decide to multi-home some sites while other satellite sites might only be single-homed to a single provider, as shown in Figure 47.

Figure 47 Enterprise Topology Dual-homed to Two MPLS VPN SPs

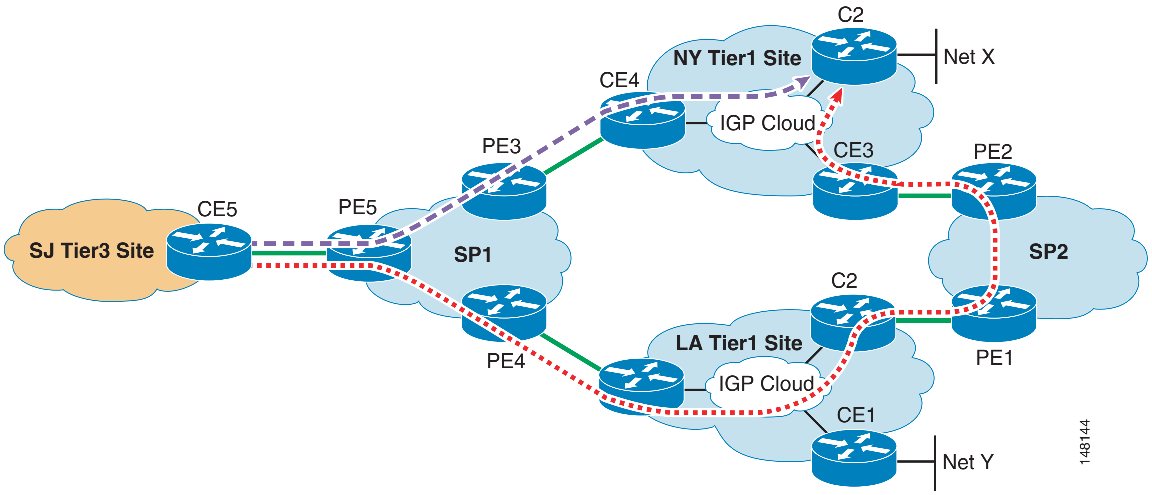


However, in case of failure, this topology may not provide the desired redundancy, and can potentially lead to loops or sub-optimal routing.

For example, assume that both SP1 and SP2 are running BGP as the PE-CE protocol with a customer. Now consider that "prefix X" is being advertised from the customer NY site to both SP1 and SP2, which in turn is sent to the LA site using eBGP. It is quite common that the customer may not do BGP with internal routers within the site. Therefore, routes (including the prefix X learned through BGP from SP1 and SP2) are redistributed into the local IGP in LA, and the local routes may be redistributed into BGP.

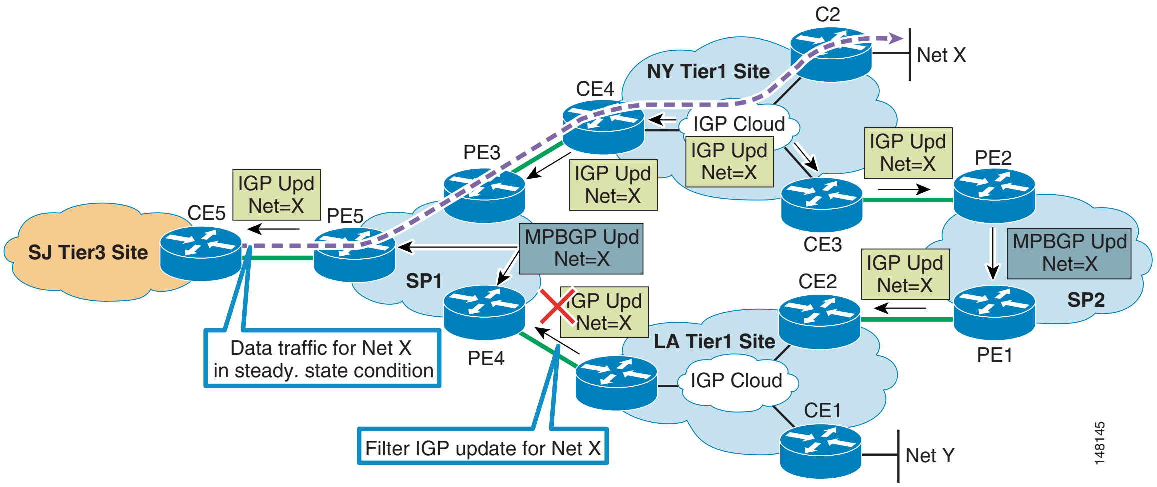
Note that because of redistribution from BGP to the local site IGP, all the BGP attributes such as AS PATH information are lost. As a result, CE1 sees the route from CE2 and vice versa through IGP, and may advertise the prefix X back to SP1 through eBGP. Depending on the SP1 topology, it is possible that the path using CE1 may be chosen as the best, and traffic from the San Jose site may transit LA and SP2 to reach prefix X. This sub-optimal route is shown by a red line in Figure 48.

Figure 48 Sub-Optimal Path to Reach Prefix X



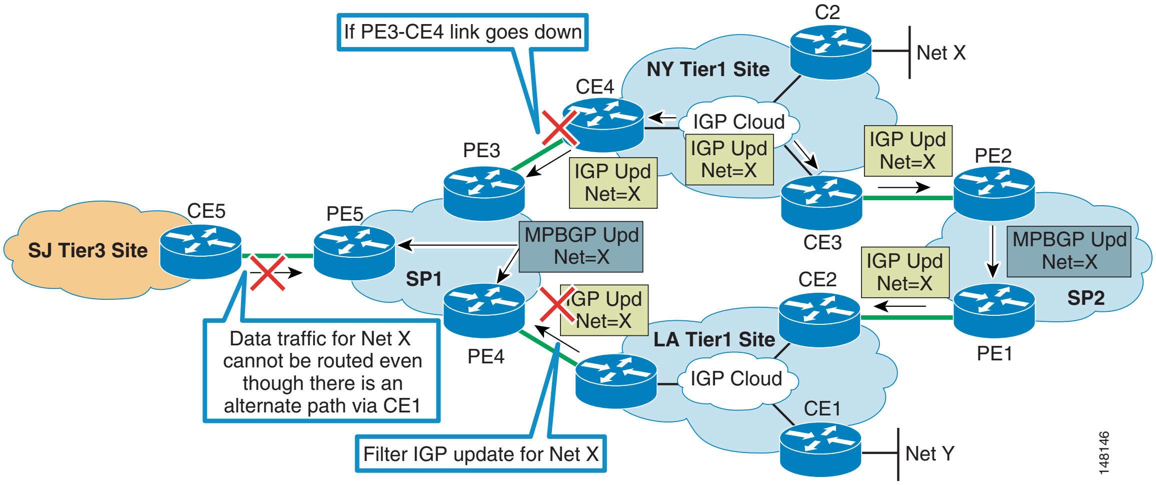
The quick solution to avoid the sub-optimal routing is to block the Net X at CE1 so that it is not advertised to SP1, as shown in Figure 49. SP1 now sees only the path through PE3 and traffic takes the desired path through PE3-CE4 to reach Net X.

Figure 49 Filtering the Update at the Customer Transit Site



Although filtering on CE1 avoids the sub-optimal routing, redundancy is lost. For example, assume that the link between PE3 and CE4 fails and the route is withdrawn from SP1. This causes all the traffic destined to prefix X from the single-homed San Jose site to be dropped, as shown in Figure 50.

Figure 50 Single-Homed Site Loses Connectivity to Dual-Homed Site After Single Failure



To avoid isolation of the single-homed site after the single failure of a link between PE3 and CE4, the better solution is to allow prefix X to be advertised from CE to SP1, and to let SP1 configure local preference or to manipulate any other BGP metrics/attributes so that PE5 prefers the path via PE3 as primary, and reverts to PE4 only if the primary path is lost.

QoS with Multihoming

It can be very challenging for an enterprise if they get two different sets of QoS offerings from different providers. If both providers can preserve the original enterprise IP DSCP marking (PIPE or Short PIPE in RFC 3270), the question of how many classes are offered by each, and how they map traffic to different classes becomes of interest. The enterprise would have to remap IP DSCP at the receiving end of the CE router link if the provider does not preserve the original enterprise IP DSCP marking (UNIFORM mode in RFC 3270).

For more information, see the following URL: http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a00800945bf.shtml.

Quality of Service Considerations

The popularity of MPLS VPNs as an alternative to private WANs introduces a number of considerations with regard to QoS. Enterprise customers can no longer achieve their historic service levels without considering SP policies. To achieve end-to-end service levels, both the customer and the SP must cooperate in their QoS designs so that they are consistent and complimentary. This section discusses the changes in QoS administration that are necessary with MPLS and suggests considerations that should be kept in mind when implementing QoS in MPLS VPN CEs. This section includes the following topics:

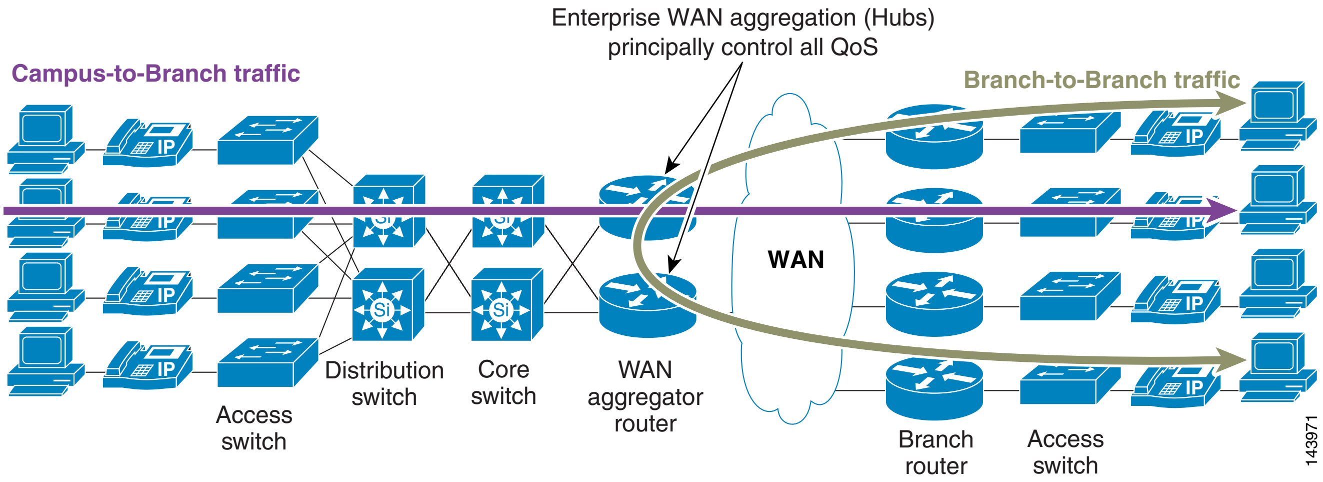
- Changes in QoS Policy Administration, page 60
- Layer 2 Access (Link-Specific) QoS Design, page 62
- Service Provider Service Level Agreements (SLA), page 62
- Enterprise-to-Service Provider Mapping Models, page 63

Note In a managed CE environment, the SP manages the QoS policies in the CE routers. However, understanding the considerations helps the customer communicate their needs to the SP efficiently and effectively.

Changes in QoS Policy Administration

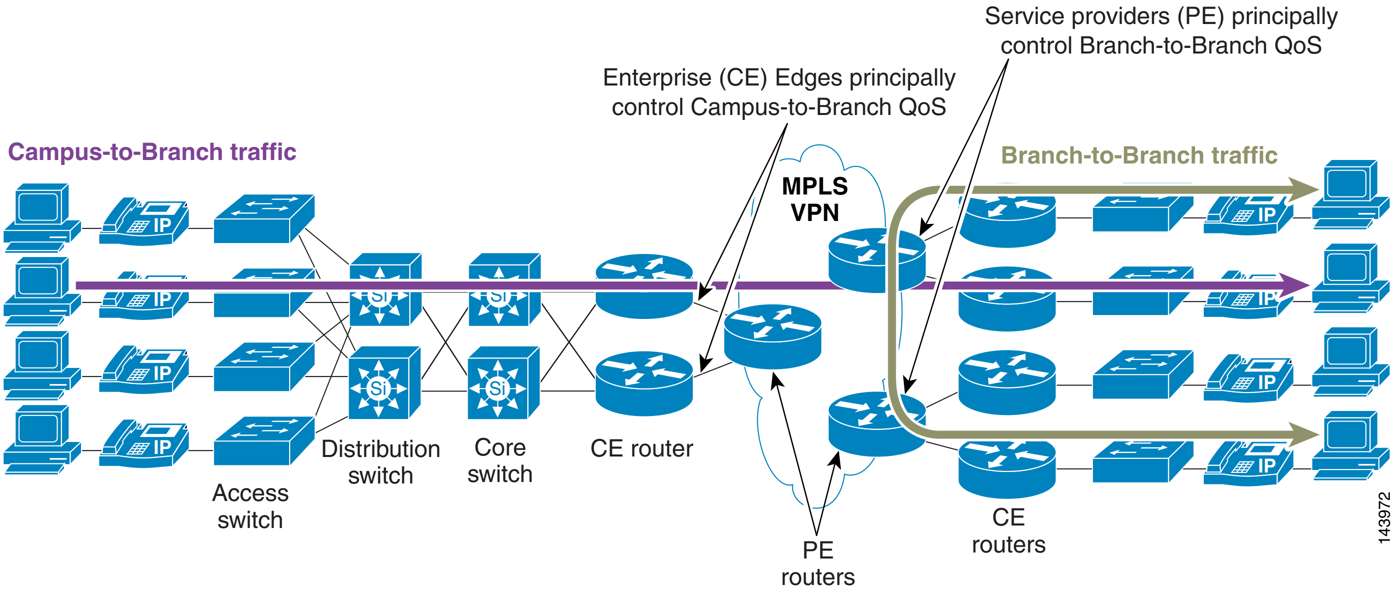
Most traditional Layer 2 WAN designs consist of a hub-and-spoke model because of cost and scalability constraints. Within such designs, QoS is administered at the hub site by the customer. Typically in a hub-and-spoke design, the hub router controls campus-to-branch traffic as well as branch-to-branch traffic (see Figure 51). Therefore, traffic received at the branches reflects the scheduling policies of the hub router.

Figure 51 QoS Administration in a Hub-and-Spoke Topology



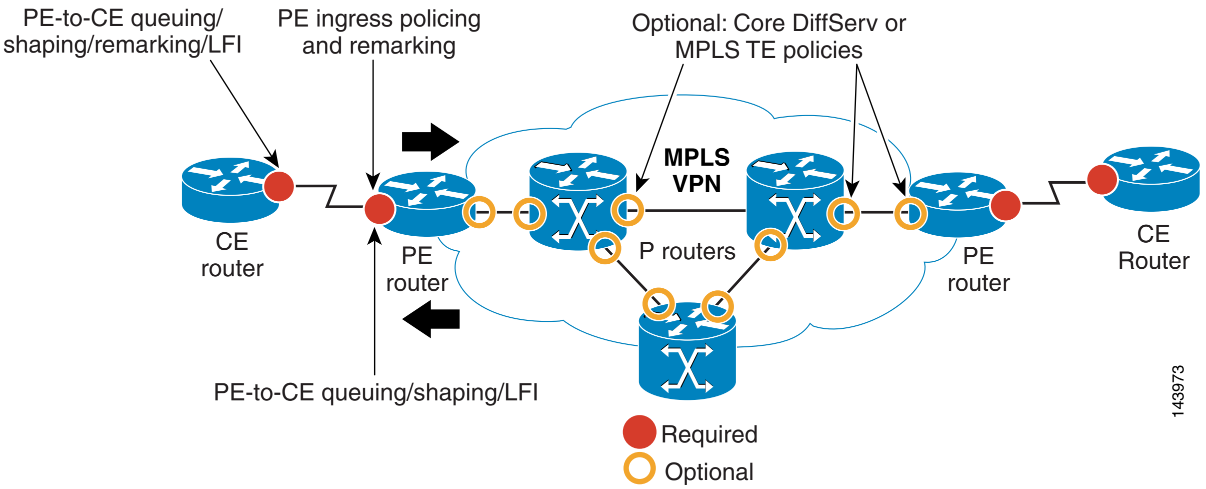
The full mesh characteristics of MPLS VPNs change the way QoS is administered. The hub router continues to administer QoS policies for campus-to-branch traffic, but it no longer fully controls the policies for branch-to-branch traffic because this traffic no longer goes through the hub router. The only way to successfully deploy needed service levels is for the SP to provision QoS policies that are compatible with the enterprise policies on all PE links to remote branches (see Figure 52).

Figure 52 QoS Administration in an MPLS VPN Depends on the SP



QoS policies can also be provisioned on the P routers within the provider network. However, providers often overprovision their MPLS core networks, so this is usually optional or not required. However, some providers might choose to implement some policies within their core MPLS network. Figure 53 shows where QoS policies can be provisioned within MPLS VPN networks.

Figure 53 Points of QoS Policy Control in an MPLS VPN Network


PE-to-CE queuing/ shaping/remarking/LFI
PE ingress policing and remarking
Optional: Core DiffServ or MPLS TE policies
MPLS VPN
CE router
PE router
P routers
PE router
CE Router
PE-to-CE queuing/shaping/LFI
143973
● Required
○ Optional

Layer 2 Access (Link-Specific) QoS Design

Migration to Layer 3 MPLS VPNs is made easier by the fact that service providers typically support Layer 2 access media such as Frame Relay and ATMs, making it unnecessary for customers to change hardware to support other access media. Because of this, Layer 2 QoS link-specific designs remain the same regardless of whether one is connecting to a Layer 2 WAN edge or Layer 3 MPLS VPN edge.

Service Provider Service Level Agreements (SLA)

It is important for customers to choose service providers that can provide the required SLAs for their converged networks. An example of such requirements for voice and interactive video is as follows:

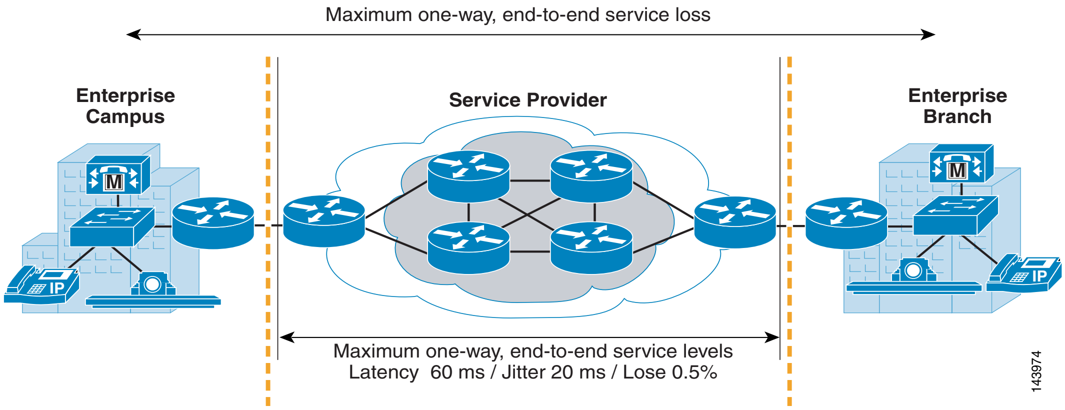
- No more than 150 ms of one-way latency from mouth to ear (per ITU G.114 standard)
- No more than 30 ms of jitter
- No more than 1 percent loss

The SP component of the SLA should be considerably tighter. The following SLAs are defined for Cisco Powered Networks (CPN) QoS Certification:

- No more than 60 ms of one-way latency from edge to edge
- No more than 20 ms jitter
- No more than 0.5 percent loss

Figure 54 shows typical SLAs for voice and interactive video.

Figure 54 Typical SLAs for Voice and Interactive Video


Maximum one-way, end-to-end service loss
Enterprise Campus
Service Provider
Enterprise Branch
M
IP
M
IP
Maximum one-way, end-to-end service levels
Latency 60 ms / Jitter 20 ms / Lose 0.5%
143974

In top text, replace everything after "…end-to-end service" with the word "loss"; that is, from "…levels" and the entire second line.

Enterprise-to-Service Provider Mapping Models

Customers and service providers must cooperate in their QoS designs to achieve the necessary end-to-end SLAs. To facilitate this cooperation, various mapping models have been developed to integrate enterprise requirements into SP solutions.

Most providers offer only a limited number of classes within their MPLS VPN clouds. This may require enterprise customers to collapse the number of classes that they use to integrate with the SP QoS model.

This section describes some of these models and caveats in regard to their implementation. It includes the following topics:

- Caveats for Integrating Enterprise and SP QoS Models, page 63
- Three-Class Provider Edge Model—CE Design, page 64
- Four-Class Provider Edge Model—CE Design, page 65
- Five-Class Provider Edge Model—CE Design, page 66

Caveats for Integrating Enterprise and SP QoS Models

The following caveats should be considered when determining how to best collapse and integrate enterprise classes into various SP QoS models.

Voice and Video

Service providers typically offer only one Real-Time class or Priority CoS. This causes a dilemma for customers deploying both voice and interactive video, each of which should be provisioned with Strict Priority treatment. Which one should be assigned to the Real-Time class? What are the implications of assigning both to the Real-Time class?

An alternative is to assign IP/VC (voice conferencing) to a non-priority class, which entails accepting the obvious caveats of lower service levels.

Note that voice and video should never be assigned low-latency queuing (LLQ) on link speeds where serialization is a factor (< 768 kbps). Packets offered to the LLQ are usually not fragmented, and large IP/VC packets can cause excessive delays for VoIP packets on slow-speed links.

Call Signaling

VoIP requires provisioning for RTP traffic, but also for Call Signaling traffic. It is important from the end users perspective that Call Signaling be protected and be given the appropriate service levels, because this directly affects the dial tone. If service providers do not offer a suitable class for Call Signaling traffic, you should consider with what other classes Call Signaling can be mixed.

On links greater than 768 kbps, Call Signaling can be provisioned into the Real-Time class, along with voice traffic. On slower speed links, Call Signaling is better assigned to one of the preferential data classes for which the SP guarantees bandwidth.

Mixing TCP and UDP

Because of the behaviors of TCP and UDP during congestion, it is generally considered a best practice to not mix TCP-based traffic with UDP-based traffic within a single SP class. Specifically, TCP transmitters throttle back flows when drops are detected, while most UDP transmitters are unaware of drops and therefore never lower their transmission rates.

When TCP flows are combined with UDP flows within a single SP class and that class experiences congestion, TCP lowers its transmission rates and potentially gives up bandwidth to UDP flow, which does not lower its transmission rate. This effect is called TCP starvation/UDP dominance. Even if WRED is enabled on the SP class, the same behavior occurs because WRED basically manages congestion for TCP-based flows.

Marking and Re-Marking

Service providers typically use Layer 3 marking attributes of packets they receive (IP Precedence or DSCP) to determine to which SP class the packets should be assigned. Therefore, the customer must mark or re-mark their traffic to be consistent with the SP admission criteria to the appropriate level of service. If such re-marking is required, it is recommended that the re-marking take place at the CE egress edge and not within the campus. This eases the management burden as service offerings change.

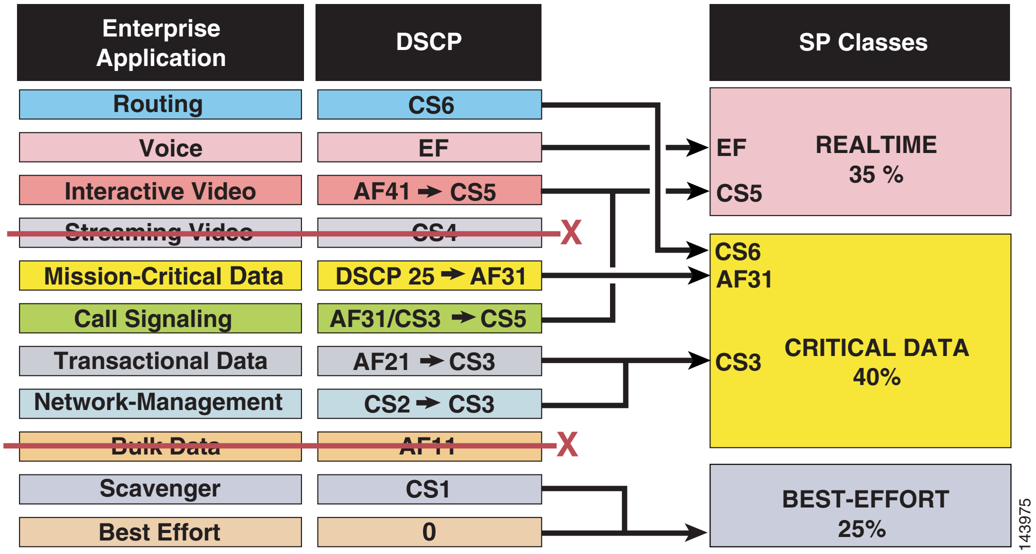
Service providers may re-mark out-of-contract traffic at Layer 3 within their cloud. This might present a problem for customers that require consistent end-to-end Layer 3 markings. In such cases, a customer can choose to re-mark traffic as it is received back from the SP MPLS VPN network, at the ingress edge of the customer CE.

Three-Class Provider Edge Model—CE Design

In this model, the SP offers three classes of service: Real-Time (Strict Priority, available in 5 percent increments), Critical Data (guaranteed bandwidth), and Best-Effort. The admission criterion for the Real-Time class is either DSCP EF or CS5; the admission criterion for Critical Data is DSCP CS6, AF31, or CS3. All other code points are re-marked to 0. Additionally, out-of-contract AF31 traffic can be marked down within the SP MPLS VPN network to AF32.

In this type of model, there is no recommended provisioning for protecting streaming video (following the guideline to not mix TCP and UDP), nor is there an SP class suitable for bulk data, which consists of large, non-bursty TCP sessions that can drown out smaller data transactions. Figure 55 shows a re-marking diagram for a three-class SP model.

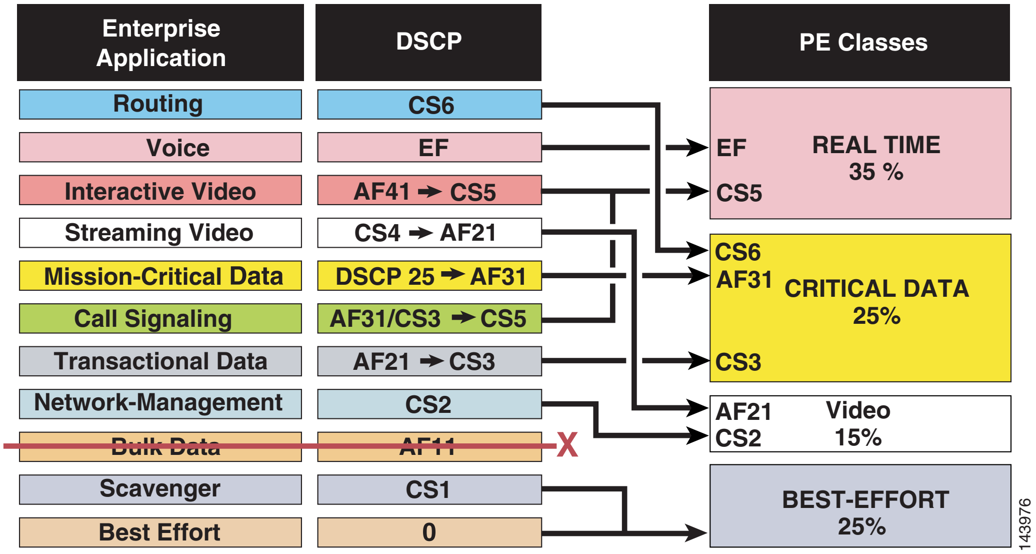
Figure 55 Three-Class Provider Edge Model



Four-Class Provider Edge Model—CE Design

Building on the previous model, a fourth class is added that can be used for either bulk data or streaming video. The admission criterion for this new class is either DSCP AF21 or CS2. The re-marking diagram shown in Figure 56 illustrates how this new class can be used for streaming video and network management traffic.

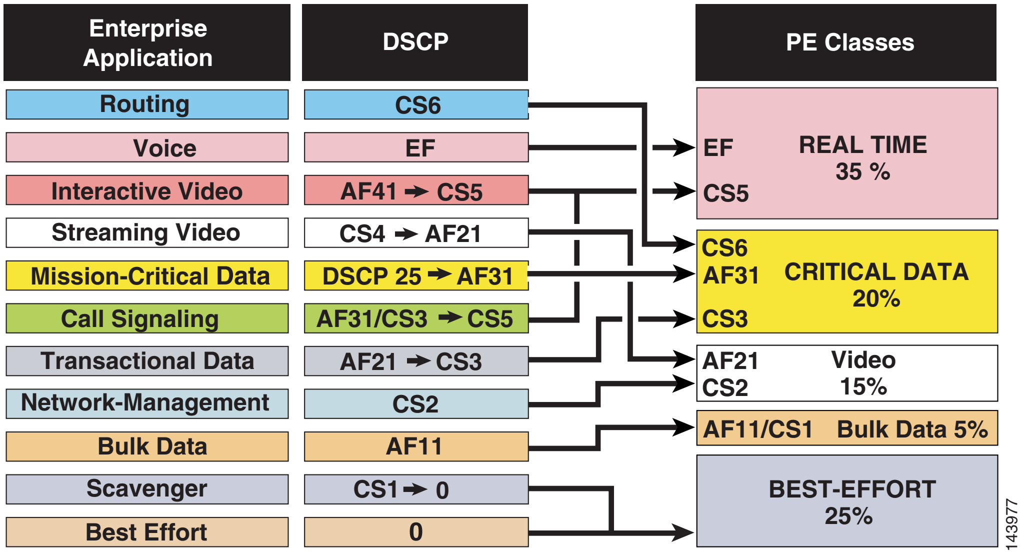
Figure 56 Four-Class Provider Edge Model



Five-Class Provider Edge Model—CE Design

Building again on the previous model, a fifth class is added that can also be used for either bulk data or streaming video, whichever was not used in the four-class model. The admission criterion for this new class is either CSCP AF11 or CS1, which necessitates the previously unrequired re-marking of the Scavenger class to DSCP 0, so that it is not admitted to the Bulk Data class but falls into the Best-Effort class. Figure 57 illustrates the re-marking required when using this new class for bulk data.

Figure 57 Five-Class Provider Edge Model



The popularity of MPLS VPNs as an alternative to private WANs introduces a number of considerations with regard to QoS. This section touched on what enterprise customers consider when planning to use an MPLS VPN service. The customer can no longer achieve their historic service levels without cooperating with the SP so that their QoS designs are consistent and complimentary of each other. QoS in an enterprise environment is discussed in much more detail in the *Enterprise QoS Solution Reference Network Design Guide* at the following URL: http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoS-SRND-Book.html.

Multicast

Multicast is widely deployed in enterprise networks because it efficiently distributes information between a single multicast source and multiple receivers. It is used to distribute services such as IPTV, video conferencing, and so on.

Because many of these enterprise customers are migrating to L3 MPLS VPNs, service providers must provide a means to allow their customers to maintain their multicast network. Because MPLS does not have the ability to natively support multicast, a separate function is required to support multicast over the MPLS VPN network. It is important for the enterprise to understand what some of these functions are and how they may impact their multicast networks.

CE-CE GRE Tunnels

A traditional function used for support of multicast over MPLS VPNs is to employ CE-CE GRE tunnels to all VPN sites participating in the multicast service. This method runs over the MPLS core, so the core does not maintain any multicast states and allows customer multicast groups to overlap. However, the use of GRE tunnels is not a scalable solution. As the number of sites requiring multicast increases, the number of tunnels necessary increases, quickly becoming difficult to manage. Although the MPLS VPN network typically supports any-to-any communication, GRE tunnels for multicast do not take advantage of this because they are typically set up in a point-to-point manner. Attempting to deploy the tunnels in an any-to-any fashion, which provides for optimal multicast routing, exponentially increases the complexity to manage the tunnels.

Multicast VPN

Multicast VPN (mVPN) was developed to enable scalable support of multicast over an MPLS environment, and is the recommended solution to service providers. When service providers provide mVPN services, enterprise customers can maintain their existing multicast configurations. Although there is minimal impact to the existing multicast network of the customer and the deployment of mVPN is the responsibility of the service provider, it is important for the customer to understand the basics of mVPN, which are discussed in this section. In addition, this section discusses issues that could affect customer use of mVPN services. For a more detailed discussion of mVPN, see Chapter 7 of *MPLS and VPN Architectures Volume II*, which is available through Cisco Press, as well as some of the following mVPN related links:

- Introduction to Multicast VPN—http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6552/prod_presentation0900aecd80310883.pdf
- mVPN Design Guide—http://www.cisco.com/en/US/tech/tk828/tech_digest09186a00801a64a3.html
- Multicast VPN Data Sheet—http://www.cisco.com/en/US/technologies/tk648/tk828/tk363/technologies_white_paper0900aecd802aea84.html

MPLS VPN versus mVPN

mVPN allows service providers to offer multicast services in a very similar manner to the way they offer MPLS VPN services for unicast. In an MPLS VPN environment, the P routers maintain routing information and labels for the global routing table only; they do not maintain any VPN information. The CE routers maintain a routing relationship with the PE router. The CE router does not peer with other CE routers but it can still, indirectly, exchange routing information with other CE routers.

The mVPN solution offered by Cisco allows service providers to offer multicast services in a similar manner. In this solution, the service provider enables native multicast in their core network. The P routers maintain multicast information for the global multicast network only; they do not maintain any VPN multicast information. The CE routers maintain multicast adjacencies (PIM adjacencies) with their local PE router. They do not peer with other CE routers, but trees can be built through the PEs to reach out to the other CE routers.

mVPN Building Blocks

The following are some of the new concepts introduced that are the basis for mVPN:

- Multicast VRFs

- Multicast domains
- Multicast distribution trees and multicast tunnel interfaces

The next sections discuss these in more detail.

Multicast VRFs

In an mVPN environment, the PE router can have each VRF associated with a multicast VRF (mVRF). The mVRF is the view of the PE router into the enterprise VPN multicast network, and contains multicast information for the VPN to which it is attached. Basically, the mVRF is the multicast routing and forwarding table of the mVRF. Multicast features and protocols such as PIM, IGMP, RP election, autoRP, and MSDP are configured in the context of the mVRF.

An mVRF is created when multicast routing is enabled for that VRF. It is identified in the same way as the underlying VRF. Multicast protocols such as IGMP and PIM are configured and operate in the context of an mVRF.

When a multicast packet is received from a particular CE device, multicast routing is done in the associated mVRF. The following summarizes some key points about mVRFs:

- Multicast VRF (mVRF)—A VRF supports both unicast and multicast forwarding tables.
- An mVRF is created when multicast routing is enabled for that VRF.
- Multicast protocols such as IGMP and PIM are configured and operate in the context of an mVRF.
- The mVRF contains only the multicast routing information for the VRFs that make up a multicast domain.

Multicast Domains

A multicast domain is a set of mVRFs that can exchange multicast information with each other. By encapsulating the original multicast packets of the customer within a provider packet using GRE, the multicast domains map all multicast groups of a customer to a single unique global multicast group in the P network. For each multicast domain supported, the service provider must supply a unique global multicast group. Because many customer multicast groups can be mapped to a single global group in the service provider network, the amount of multicast state information that P routers must hold is deterministic and is not dependent on the customer multicast deployment.

A PE router supporting an mVPN for a customer is part of the multicast domain of that customer. PE routers can support mVPNs for more than one customer; therefore, the PE router can be a member of many multicast domains.

Multicast Domain Tree

Multicast distribution trees (MDTs) are multicast tunnels through the P network that encapsulate and transport customer multicast traffic within the corresponding multicast domain through the provider network. Note that the GRE packets are IP switched through the core. There are no MPLS labels applied to MDT traffic.

There are the following two types of MDTs:

- Default MDT

  This is a distribution tree created by the mVPN configuration. When configured, it is always on. This MDT is used by the mVRF to send multicast control traffic between PE routers in a common multicast domain as well as sending low-bandwidth multicast traffic.

  When a VRF is multicast enabled, it must also be associated with a default MDT. After multicast is enabled on a VRF and a default MDT is configured, the PE router joins the default MDT for that domain, whether or not there are any sources or receivers that are active.

Currently, an mVRF can belong to only one default MDT, and therefore extranets cannot be formed between mVPNs. This functionality is expected to be available soon.

- Data MDT

  This is a distribution tree that is created on demand to minimize the flooding by sending data only to PE routers that have active multicast receivers. The data MDT is used for data traffic only. Control traffic continues to use the default MDT.

  As previously mentioned, data MDTs are created automatically when data traffic on the default MDT exceeds a configured threshold. Only PE routers that have interested receivers join the data MDT.

Note The data MDT capability is an optional capability and must be configured in the PE router.

Multicast Tunnel Interfaces

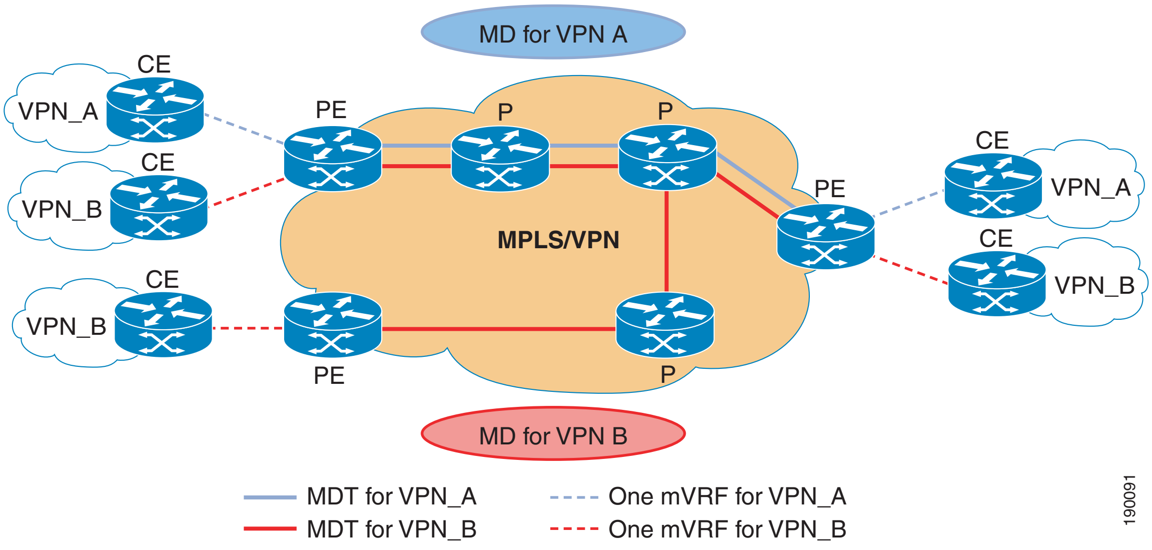
When the default MDT is configured, a multicast tunnel interface (MTI) is created dynamically. This interface is not configured manually. It takes on the interface characteristics of another interface. The MTI is the interface that connects the customer multicast environment (mVRF) to the service provider global environment (MDT). One MTI is created for each mVRF, and the same MTI is used to forward traffic whether it is to the default MDT or the data MDT. The PE routers that are part of the same multicast domain form PIM adjacencies with each other via the MTI.

Unicast packets are not forwarded over the MTI because this interface does not appear in the unicast routing table of the associated VRF.

Customer multicast packets that are sent to the MTI are encapsulated into a provider packet, using GRE, and sent along the MDT.

Figure 58 illustrates the concepts discussed above and where they fall within the multicast network.

Figure 58 mVPN Concepts



Customer Considerations

As mentioned previously, the mVPN solution provided by Cisco allows the enterprise customer to transparently interconnect its private multicast network across the provider network. The customer does not have to change how their enterprise multicast network is configured. On the customer side, all standard customer topologies using auto-RP, BSR, Static RP, or PIM SM, PIM SSM work fine. This section discusses what might not work in an mVPN environment, and provides design guidance for auto-RP customer setups.

Triggering Data MDTs

As mentioned above, data MDTs are multicast distribution trees that are created on demand to minimize flooding of multicast packets. If the mVPN provider provides data MDTs, the enterprise customer should avoid using PIM BiDir mode to take advantage of the data MDT. If running PIM SM, the customer should avoid setting the spt-threshold to infinity.

This recommendation is related to the multicast states created in the PE by the PIM BiDir and PIM SM protocols as well as to how the data MDT is triggered. When a traffic threshold is exceeded on the default MDT, the PE router that is connected to the mVPN source of the multicast traffic can switch the (S,G) from the default MDT to a group associated with the data MDT. However, the data MDT is triggered only by an (S,G) entry in the mVRF. A (*,G) entry does not trigger the data MDT.

With PIM BiDir, there are no source tree (S,G) entries in any of the multicast routers because all multicast routing entries for bidirectional groups are on a shared tree (*,G). Therefore, the default MDT is used for all multicast traffic, regardless of the bandwidth.

With PIM SM, multicast traffic initially uses a shared tree (*,G). When the traffic on the shared tree reaches a certain threshold (spt threshold), the last hop router can join a shortest path tree to the source (S,G). If the spt-threshold is set to infinity, (S,G) trees are never created and the data MDT is not triggered.

There is an exception to the above: with PIM SM, there is always an (S,G) state between the rendezvous point (RP) and the multicast source and all routers in between, even if the spt-threshold is set to infinity. Therefore, if the PE router is the RP or is between the source and the RP, the data MDT can be triggered for the tree rooted on the customer RP.

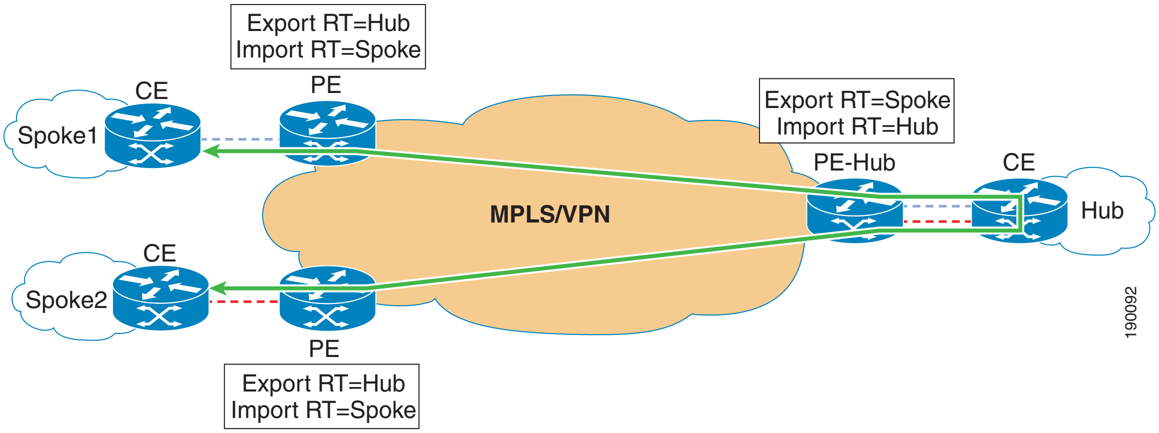
MPLS VPN Hub and Spoke Topologies

In certain circumstances, it may be desirable to use a hub-and-spoke topology within an MPLS VPN environment. This may be because of the need to have certain services centralized at a hub, or the VPN customer may have a policy requiring all connectivity between its sites to be through a hub. However, mVPN does not work over this topology. To understand why, it is necessary to know how a hub-and-spoke topology may be implemented in an MPLS VPN environment.

Note The hub-and-spoke topology described here is a simple topology used for the purpose of understanding the implementation. Other variations of the topology that can be implemented are not be discussed here.

In a hub-and-spoke topology, the spoke sites export their routes to the hub site. The hub site then re-exports the spoke site routes through a second interface (physical or logical) using a different route target, so that the other spoke can import the routes. This is explained a bit more in Figure 59 and the following text.

Figure 59 Hub-and-Spoke Topology



In the hub-and-spoke topology shown in Figure 59, when Spoke1 and Spoke2 communicate with each other, they do so via the hub site. For this to happen, the hub site requires two connections to the PE router (PE-Hub): one that imports all spoke routes into the hub and another that is used to export hub routes back to the spokes. Each of these connections has its own VRF, denoted by the blue and red lines. The spoke PE routers export their routes using a route target of Hub. The PE-Hub router imports these routes into one of the VRFs (for example, VRF Red). These spoke routes are then advertised to the hub over the link associated with VRF Red. At some point, the hub site advertises these routes back to the PE-Hub over the link associated with VRF Blue. The PE-Hub router then exports these routes with a route target of Spoke when advertised to the spoke PE routers. The spoke PE router imports these routes that have a next hop of the PE hub router.

The fact that the PE-Hub router must use more than one VRF to support this topology means that it also needs to use more than one mVRF to support mVPN in this topology. Currently, however, each mVRF can be a member of only one VPN domain. Multicast joins and prunes cannot cross from one mVRF to another and therefore, mVPN does not work in this environment.

Using Auto-RP

When a customer is using auto-RP, the interfaces of all the routers must be configured for PIM sparse-dense mode. With auto-RP, the RP announces itself by flooding its advertisements, via PIM dense mode, throughout the multicast network. PIM dense mode is used only for the two auto-RP groups (224.0.1.39 and 224.0.1.40). The disadvantage to configuring sparse-dense mode on the interfaces is that if the RP is ever lost, and the interfaces are in dense mode, all traffic starts being dense mode flooded. As an alternative, customers can configure **ip PIM auto-rp listener**. This enables the interfaces to listen to dense mode flooding only for the auto-RP groups. The interface is in sparse mode, so there is no risk of flooding if an RP is lost.

Summary

To summarize, the following are the benefits of using a service provider that has deployed mVPN:

- There is no need for the customer to employ CE-CE GRE tunnels.
- Multicast configuration changes are not required in customer networks.
- The existing customer multicast deployment, including PIM modes and RP placement/discovery mechanism, is unaffected, with the exception mentioned about triggering the data MDT.
- Ability to have worldwide remote multicast networks connected via the same mVPN service.

Security

This section discusses three levels of security:

- General router security, which can be applied to any single router in the network
- Securing the network or links, specifically focused on securing the PE-CE connectivity
- Securing the data that is sent over an MPLS VPN

General Router Security

Router security is an important part of securing any network infrastructure. It is not specific to an L3 MPLS VPN environment but is included here for completeness.

This section describes measures that can be taken to secure a single router. These measures can then be applied to the rest of the routers in the network. Whether you are an enterprise customer or a service provider, network security begins in an MPLS VPN environment with the physical security of the networking equipment. This means reducing the risk of any physical, environmental, or electrical damage. The following steps protect the equipment:

- Limit physical access to the router to only authorized personnel.
- Log all entry attempts to the equipment.
- Reduce the risk of environmental threats by locating the equipment in an environmentally friendly location that offers temperature control, plenty of airflow, and low humidity.
- Use redundant power supplies and/or backup power sources to ensure the availability of the equipment.
- Use UPS systems and power conditioning equipment to avoid voltage spikes and brownouts.

Securing Access to the Routers

After controlling physical access to the router, control access to the various lines that offer interactive access to the router: the console, virtual terminal (VTY), asynchronous lines (TTY), and the auxiliary line (AUX). Anyone that can access a line and log onto a router can manipulate the router in malicious ways. Controlling access to the lines and controlling logins helps to prevent any inappropriate use of the router.

Access control can be accomplished locally on the router by using such methods as local username authentication and line password authentication, or by limiting which sources can access the specific lines. The following shows a configuration example that allows Telnet from a specific source only, and logs any Telnet attempts from other sources.

```
Line vty 0 4
    login
    password h0ttama1e

    access-class 13 in
    #applies access-list 13 to inbound telnet requests

    ip access-list 13 permit 172.16.1.1
    ip access-list 13 deny any log
    #creates an ACL that permit access from host 172.16.1.1 while
    #denying and logging attempts from any other sources.
    transport input telnet
    #only incoming telnet sessions are permitted
```

Note When applying an access class to a set of VTYs in an MPLS VPN environment, the parameter “vrf-also” is required; otherwise, all Telnet attempts to the router are refused.

In the above example, the protocol used to interact with the router is limited to Telnet on the VTY lines. Interactive access to a router can also be accomplished using Secure Shell (SSH). The security and encryption inherent in SSH make it a good choice for an access protocol. A router can be enabled as an SSH server allowing SSH clients to make a secure, encrypted connection to the router. This connection provides similar functionality to that of an inbound Telnet connection.

Controlling access locally as shown above does not offer the same degree of access control that is possible by using Authentication, Authorization, and Accounting (AAA). AAA network security services provide the primary framework through which you can set up access control on your router. It allows you to control who is allowed access to the router and what services they are allowed to use when they have access.

For detailed instructions and examples for configuring AAA, SSH, and other security features, see the *Cisco IOS Security Configuration Guide* at the following URL: http://www.cisco.com/en/US/docs/ios/12_3/featlist/sec_vcg.html.

Disabling Unnecessary Servers and Services

Many of the built-in services in Cisco IOS are not needed and can represent a security risk when not being used for a specific reason. These services should be disabled unless there is an explicit reason to have them enabled. Some of these are off by default but this can be dependent on the Cisco IOS version. Table 5 provides a list of some of these servers and services, some of which are important and should be turned off only if not used.

Table 5 Built-in Services

Feature	Description
Cisco Discovery Protocol (CDP)	Proprietary Layer 2 protocol between Cisco devices
TCP/UDP small servers	Standard services with port numbers below 10, such as echo, discard
Finger	Listens for “finger” requests (user lookup service) from remote hosts
HTTP server	Used by Cisco IOS device offering web-based configuration
BOOTP server	Allows other routers to boot using operating system from this (server) router
Configuration auto-loading	Router attempts to load it’s configuration via a TFTP server
IP source routing	Allows sender of the packet to specify the route the packet should take towards the destination
Proxy ARP	Router responds to ARP requests on behalf of another device
IP directed broadcast	Packets can identify a target LAN for broadcasts
IP unreachable notifications	Router explicitly notifies senders of incorrect IP addresses

Table 5 Built-in Services (continued)

IP mask reply	Router sends an interfaces IP address mask in response to an ICMP mask request
IP redirects	Router sends an ICMP redirect message instructing an end node to use a specific router as its path to a destination
NTP service	Router acts as a time server for other devices and hosts

As mentioned, if the servers or services are not needed, disabling them reduces the security risks they represent when enabled but not in use. When they are in use, limiting access to them can be done with the use of access control lists (ACLs). For an explanation of how ACLs can filter network traffic, see *Configuring IP Access Lists* at the following URL: http://www.cisco.com/en/US/products/sw/secursw/ps1018/products_tech_note09186a00800a5b9a.shtml.

Securing the PE-CE Connectivity

The PE and CE routers represent the network edge in an MPLS VPN network. Because the PE-CE interconnection involves two separate corporate entities, natural concerns arise regarding security and stability of the interconnection. Specifically, at Layer 3, the PE or the CE can be flooded with unwanted routes from its neighbor. Instability in the routing environment can indirectly affect the routing environment of the neighbor. An attacker can inject invalid routes into either network, forcing traffic to take unsecured paths. This section focuses on CE-PE routing security practices that can alleviate some of the above concerns.

The service provider usually chooses which routing protocol to use on the PE-CE link and how to secure it, but understanding options the provider has is important to the customer.

Static Routing

Regarding the PE-CE routing protocol, static routing offers both the enterprise and the service provider the most stable and controlled PE-CE environment. There is less CPU impact because static routing does not require any dynamic process to be running between the PE and CE routers. You do not have to worry about an invalid peer relationship being created. Static routes require that the routes be entered directly into the configurations of the CE and PE routers, which lessens the chance that invalid routes will be injected into the routing environment. Obviously, the drawback to static routing is that each time the network grows or changes in any way, manual configuration is required.

Dynamic Routing

Dynamic routing protocols are used more often because they can dynamically manage changes in the routing environment after the initial configuration is entered. Multiple routing protocols can be used between the PE and CE routers; the choice is usually determined by which protocol the service provider supports. In most cases, eBGP is the preferred protocol for service providers based on its stability, scalability, and control features. BGP and other protocols that can be used as PE-CE protocols have been discussed in this document already. This section discusses making the PE-CE routing protocols more secure.

PE and CE routers can be connected via shared media, such as Ethernet, or may be misconnected in a way that allows for invalid peers to be introduced. To prevent routes from being accepted from non-authorized neighbors, Cisco recommends that MD5 authentication be used between the PE and CE routers.

Note For examples of the various protocols that are sometimes used as PE-CE protocols, and how MD5 authentication is used for each, see "Neighbor Router Authentication: Overview and Guidelines" at the following URL:
http://www.cisco.com/en/US/products/sw/iosswrel/ps1835/products_configuration_guide_chapter09186a00800ca7d2.html

Preventing Excessive Route Flooding

Within an MPLS VPN environment, the CE and/or the PE can be exposed to the intentional or unintentional flooding of an excessive number of routes. This act can potentially lead to a depletion of router resources that causes the router to stop processing packets or to reload. BGP, if used as the PE-CE protocol, allows the user to define a maximum number of routes that are accepted from a neighbor. When the maximum has been reached, the offending neighbor is warned and eventually restarted. The following is a sample BGP configuration that defines the maximum prefix limits.

```
router bgp 131
no synchronization
no auto-summary

address-family ipv4 vrf red
neighbor 141.1.250.2 remote-as 250
neighbor 141.1.250.2 activate
neighbor 141.1.250.2 maximum-prefix 45 80 restart 2
(the above command defines a 45 prefix maximum for routes delivered from this neighbor,
with a warning message issued at 80% of 45, and the BGP session restarted on exceeding 45,
with a 2 minute interval for retry)
no auto-summary
no synchronization
exit-address family
```

In a managed VPN service where the CE router is managed by the service provider, the PE-CE security is provided by the service provider. In an unmanaged VPN environment, where the enterprise customer owns and controls the CE router, PE-CE security must be a collaboration between the enterprise customer and the service provider.

Securing the Data over an MPLS VPN Network

MPLS VPNs and IPsec VPNs have been compared and contrasted with each other to determine which VPN technology is better or more suitable for a given environment. Both technologies have their benefits, depending on the scenario. MPLS VPNs allow for an architecture that can provide any-to-any data paths between VPN sites, which service providers can offer to customers at lower prices. However, MPLS VPNs do not provide data encryption. IPsec VPNs mainly benefit the security of the customer network because data can be encrypted and authenticated, and the integrity can be maintained. IPsec and MPLS VPNs can be deployed together to enhance the VPN architecture.

In an MPLS VPN environment, IPsec can overlay secure tunnels between two peers or endpoints. When implementing IPsec in any network, deciding where the IPsec endpoints should be applied and how the tunnels should be established are two decisions that need to be made.

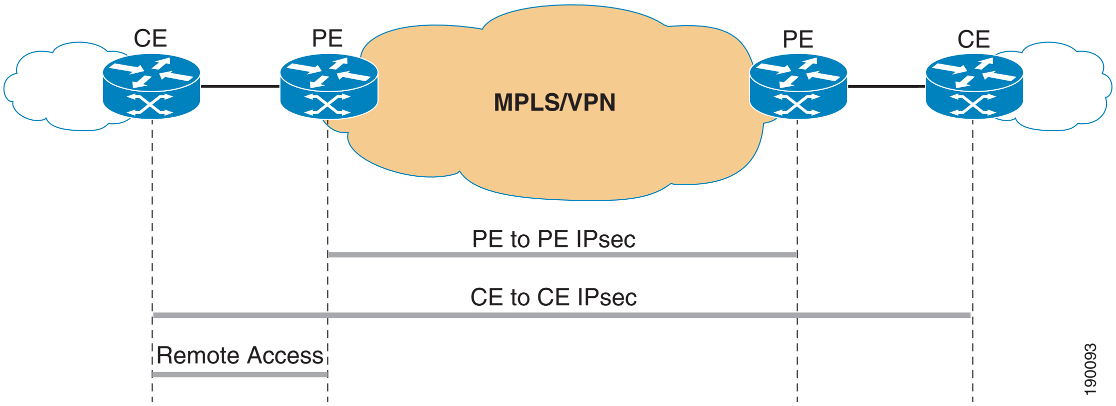
End Point Selection

In an MPLS VPN network, the IPsec endpoints can be established as the following:

- Between the CE routers of the VPN
- Between the PE router of the VPN
- Between a point in the VPN and a PE router

These are shown in Figure 60.

Figure 60 IPsec Endpoints



In a self CE-managed environment, CE-CE IPsec can be controlled and configured by the enterprise customer and is discussed in more detail in the following section.

CE-CE IPsec

In an MPLS VPN environment, the infrastructure that interconnects the CE devices is the MPLS core provided by the service provider. In subscribing to the MPLS VPN service, the customer is essentially “trusting” that the service provider is providing adequate security through that infrastructure. If for any reason the customer determines that infrastructure is to be “untrusted”, this is a reason to implement CE-CE IPsec. For example, the service provider might inadvertently make the VPN insecure by misconfiguring a PE router. If the enterprise wants to protect itself from such possibilities, implementing CE-CE IPsec is an option.

In other cases, enterprises may have a security policy that requires them to encrypt traffic before handing it off to any third party, which can be accomplished using CE-CE IPsec.

When CE-CE IPsec is implemented, it performs the following anywhere between the CEs:

- Protects against eavesdropping, keeping the data confidential
- Protects against the insertion of any invalid packets because the source of the packets is authenticated
- Authenticates the packets sent by the IPsec sender, ensuring the integrity of the packets
- Protects against the replay of old or duplicated packets (anti-replay)

Configuring and deploying IPsec has been discussed in many other venues, so this document does not discuss it in detail. For more information on IPsec, see *Configuring Security for VPNs with IPsec* at the following URL:
http://www.cisco.com/en/US/docs/ios/security/configuration/guide/sec_cfg_vpn_ipsec.html.

IPsec Scalability Issues

As mentioned above, IPsec encrypts traffic between two endpoints (or peers). The encryption is done using a shared “secret”. Because the sharing is only between the two endpoints, the encrypted network is basically a collection of point-to-point links. Scaling these point-to-point links usually means organizing the network in a hub-and-spoke network. An MPLS VPN infrastructure offers any-to-any connectivity. Trying to use IPsec and still use the any-to-any model quickly becomes a configuration management nightmare. Each time a CE is added to the network, each CE needs to be configured to peer with that CE. Even in a hub-and-spoke environment, where spokes communicate only with each other via the hub, the hub configuration is difficult to manage.

Some of the scalability issues result from the complexity in establishing the IPsec tunnels. Newer deployment models improve the scalability. Options for establishing IPsec tunnels are as follows:

- Static IPsec—Every IPsec node is configured statically with all its IPsec peers, authentication information, and security policies. This is the least scalable of all.
- Dynamic IPsec—More scalable than the previous, this model allows the hub in hub-and-spoke environments to be configured without specific information for each spoke. The tunnel is established when the spoke tries to initiate an IPsec security association with the hub.
- Dynamic Multipoint VPN (DMVPN)—This is a very scalable way to dynamically establish IPsec tunnels on demand and is subsequently discussed in more detail.

Other scalability issues with IPsec are as follows:

- IPsec uses an ACL to define what data is to be encrypted. Each time a new network is added behind a spoke or hub, the ACL on both the hub and spoke routers must be changed so that traffic from the new network is encrypted.
- IPsec does not support the encryption of broadcast or multicast packets, which means it does not support dynamic routing protocols. This is solved by using GRE tunnels in combination with IPsec encryption. GRE + IPsec must know the endpoint peer address. The IP addresses of the spoke are, in many cases, connected directly to the Internet via their ISP. The address of the interface connected to the Internet need to be fixed, and therefore can not be DHCP assigned.
- If spokes need to communicate with each other directly, the network must become fully meshed. This might require that the spoke routers be more powerful to handle the additional load caused by establishing several IPsec peers.

Dynamic Multipoint IPsec VPNs (DMVPNs) is a Cisco solution that uses multipoint GRE (mGRE) and Next Hop Resolution Protocol (NHRP) with IPsec to solve the above problems.

Benefits of DMVPN

Hub Router Configuration Reduction

Currently, for each spoke router, there is a separate block of configuration lines on the hub router that define the crypto map characteristics, the crypto access list, and the GRE tunnel interface. DMVPN allows users to configure a single mGRE tunnel interface, a single IPsec profile, and no crypto access lists on the hub router to handle all spoke routers. The size of the configuration on the hub router remains constant even if spoke routers are added to the network.

Automatic IPsec Encryption Initiation

Without DMVPN, the IPsec encryption tunnel is not initiated until there is data traffic that requires the use of the IPsec tunnel. Packets are dropped during the time it takes to complete the initiation. With DMVPN, IPsec is started immediately for point-to-point and mGRE tunnels.

Dynamic Tunnel Creation for Spoke-to-Hub Links

With DMVPN, there is no need to configure the hub router with GRE or IPsec information about the spoke routers. The spoke router is configured with NHRP information about the hub router and automatically notifies the hub router of its current physical interface IP address by using NHRP. With DMVPN, one router is the hub and all other routers, the spoke routers, are configured with tunnels to the hub. These spoke-to-hub tunnels are up continuously, and the spokes do not need configuration for direct tunnels to any other spokes.

Dynamic Tunnel Creation for Spoke-to-Spoke Traffic

As mentioned previously, the only tunnels that are up continuously are the spoke-to-hub tunnels initiated by the spokes. When the spokes want to send packets to another spoke or network behind the spoke, they use NHRP to resolve the destination of the next hop spoke.

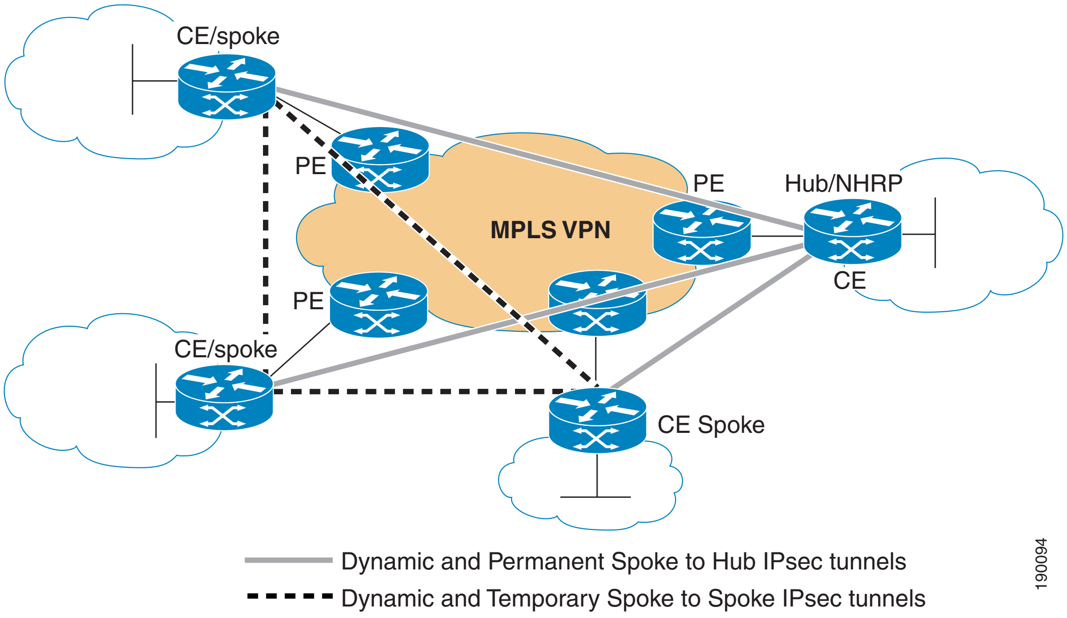
The hub acts as the NHRP server handling the request from the spokes. The two spokes then dynamically create an IPsec tunnel between them. After a period of inactivity, the tunnel is torn down.

The spoke-to-spoke links are established on demand whenever there is traffic between spokes. Packets can then bypass the hub and use the spoke-to-spoke tunnel.

DMVPN in an MPLS Environment

When using DMVPN in an MPLS VPN environment, one of the CE routers acts as the hub. The other CE routers act as spokes to establish the permanent spoke-to-hub tunnels, as shown in Figure 61.

Figure 61 Using DMVPN in an MPLS Environment

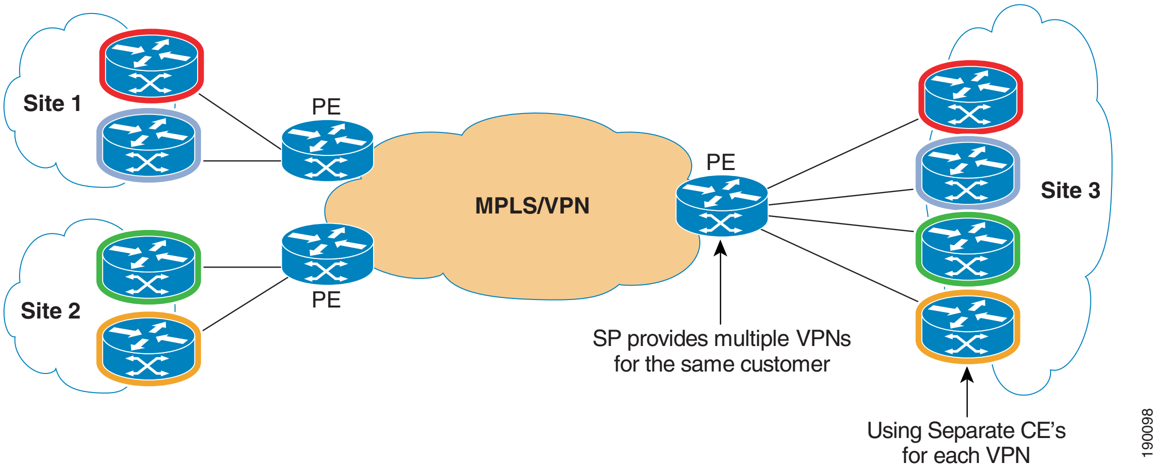

CE/spoke
PE
PE
Hub/NHRP
MPLS VPN
CE
PE
CE/spoke
CE Spoke
190094
Dynamic and Permanent Spoke to Hub IPsec tunnels
Dynamic and Temporary Spoke to Spoke IPsec tunnels

Using Multi-VRFs

In some cases, enterprise customers may want to introduce a level of separation between entities within their own network; for example, between the research and development and marketing departments, or trading and retail departments. To separate the organization in such a way means the company must divide its intranet VPN into multiple independent departmental VPNs.

One way to implement this separation is to have the MPLS VPN service provider provide multiple VPNs to the same enterprise customer. The customer, or the provider in the case of a managed CE service, can then deploy multiple CE routers at each site. Each CE router then serves one of the VPNs for the enterprise, as shown in Figure 62.

Figure 62 Multi-VRF—Example 1


Site 1
PE
MPLS/VPN
PE
Site 3
Site 2
PE
SP provides multiple VPNs for the same customer
Using Separate CE's for each VPN
190098

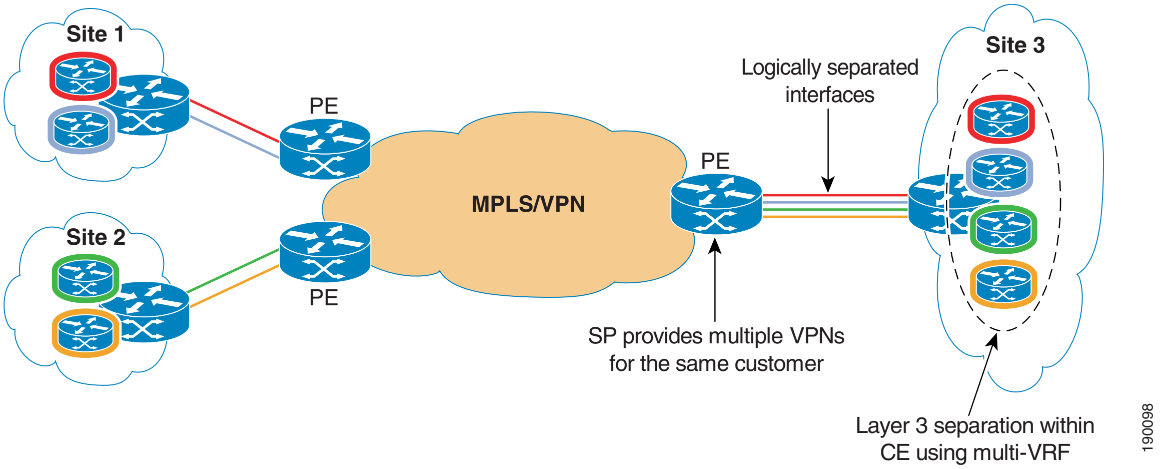
This scenario has an additional cost because there are multiple CE routers required, and multiple links attaching to the MPLS VPN backbone.

A more efficient and cost effective way to implement this separation is to move some of the MPLS VPN functionality into the CE. Specifically, the VRF functionality allowing the CE router to have multiple independent routing entities can be added to the CE.

Each routing entity then has its own address space, routing table, and routing process.

Figure 63 shows this implementation.

Figure 63 Multi-VRF—Example 2


Site 1
Site 2
Site 3
PE
PE
PE
MPLS/VPN
Logically separated interfaces
SP provides multiple VPNs for the same customer
Layer 3 separation within CE using multi-VRF
190098

The advantage in this implementation is lower cost; no additional CEs are required to reach the desired separation.

To reduce the cost even more, instead of provisioning multiple links from the new CE to the PE, you can use Frame Relay on the point-to-point serial links. Sub-interfaces can then be used to logically separate the links. This is also shown in Figure 63.

If the PE-CE connection is an Ethernet-type technology, VLANs can be implemented.

Using this technology, which is known as multi-VRF functionality, you can have multiple VRFs within a single CE router, allowing you to logically separate VPNs. Bringing this functionality into the CE avoids the need for deploying multiple CEs, or the need for deploying a PE router at the customer site.

For more detailed examples and explanations of multi-VRFs and how they are used, see Chapter 4 of the Cisco Press book *MPLS and VPN Architectures Vol. II*.

Summary

Layer 3 MPLS VPNs offer customers an effective way to expand their networks geographically while establishing any-to-any connectivity and lowering costs by replacing dedicated circuits such as Frame Relay or ATM. Choosing a provider for an MPLS VPN service and integrating with that provider can be a daunting task. Taking the time to assess your business requirements, environment, and objectives ensures your success when selecting and integrating with a service provider.

This guide has introduced some criteria to consider when selecting a provider and has provided some general guidelines for integration with a L3 MPLS VPN service.

References

- Cisco Powered Network—Find recommended service providers

  http://www.cisco.com/cpn

- *Enterprise QoS SRND*

  http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoS-SRND-Book.html

- *8 Questions to ask when choosing a service provider*

  http://www.cisco.com/cdc_content_elements/flash/cpn/flash.html
- *The Move to MPLS-Based VPNS: Exploring Service Options*

  http://www.cisco.com/en/US/solutions/collateral/ns341/ns524/ns546/ns193/net_implementation_white_paper0900aecd800f6d9a.html