Advanced Applications Over MPLS-Based VPNs

Cisco IP Communications

This section highlights and describes the design differences and requirements of integrating Cisco IP Communications with the self-managed MPLS MAN. It is not intended to provide full details of general enterprise IP Communications design, which is highly complex.

Much of the content in this section that specifically applies to the self-managed MPLS MAN has been taken from the Cisco Enterprise IP Telephony SRND for CallManager 4.1 and the Enterprise QoS Solution Reference Network Design Guide.

See the following URLs for complete details of general deployment design guidance for IP Communications:

- Enterprise IP telephony SRND—http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html
- QoS SRND — http://www.cisco.com/univercd/cc/td/doc/solution/esm/qossrnd.pdf

Overview of Cisco IP Communications Solutions

Cisco IP Communications solutions deliver fully integrated communications by enabling data, voice, and video to be transmitted over a single network infrastructure using standards-based IP. Leveraging the framework provided by Cisco IP hardware and software products, Cisco IP Communications solutions deliver unparalleled performance and capabilities to address current and emerging communications needs in the enterprise environment. Cisco IP Communications solutions are designed to optimize feature functionality, reduce configuration and maintenance requirements, and provide interoperability with a wide variety of other applications. Cisco IP Communications solutions provide this capability while maintaining a high level of availability, QoS, and security for the network.

Cisco IP Communications encompass the following solutions:

- IP telephony—Transmits voice communications over the network using IP standards. The Cisco IP Telephony solution includes a wide array of hardware and software products such as call processing agents, IP phones, video devices, and special applications.
- Unified communications—Delivers powerful unified messaging (email, voice, and fax messages managed from a single inbox) and intelligent voice messaging (full-featured voicemail providing advanced capabilities) to improve communications, boost productivity, and enhance customer

service capabilities across an organization. Cisco Unified Communications solutions also enable users to streamline communication processes through the use of features such as rules-based call routing, simplified contact management, and speech recognition.

- Rich-media conferencing—Enhances the virtual meeting environment with a integrated set of IP-based tools for voice, video, and Web conferencing.
- Video telephony— Enables real-time video communications and collaboration using the same IP network and call processing agent as the Cisco IP Telephony solution. With Cisco Video Telephony, making a video call is now as easy as dialing a phone number.
- Customer contact—Combines strategy and architecture to promote efficient and effective customer communications across a globally-capable network by enabling organizations to draw from a broader range of resources to service customers, including access to an unlimited pool of agents and multiple channels of communication as well as customer self-help tools.
- Third-party applications—Cisco works with leading-edge companies to provide the broadest selection of innovative third-party IP telephony applications and products focused on critical business needs such messaging, customer care, and workforce optimization.

This section describes the deployment of the Cisco IP Telephony solution for a self-managed MPLS MAN network. This topology is provisioned with separate VRFs for voice and data devices. Inter-VPN voice communication occurs over PSTN.

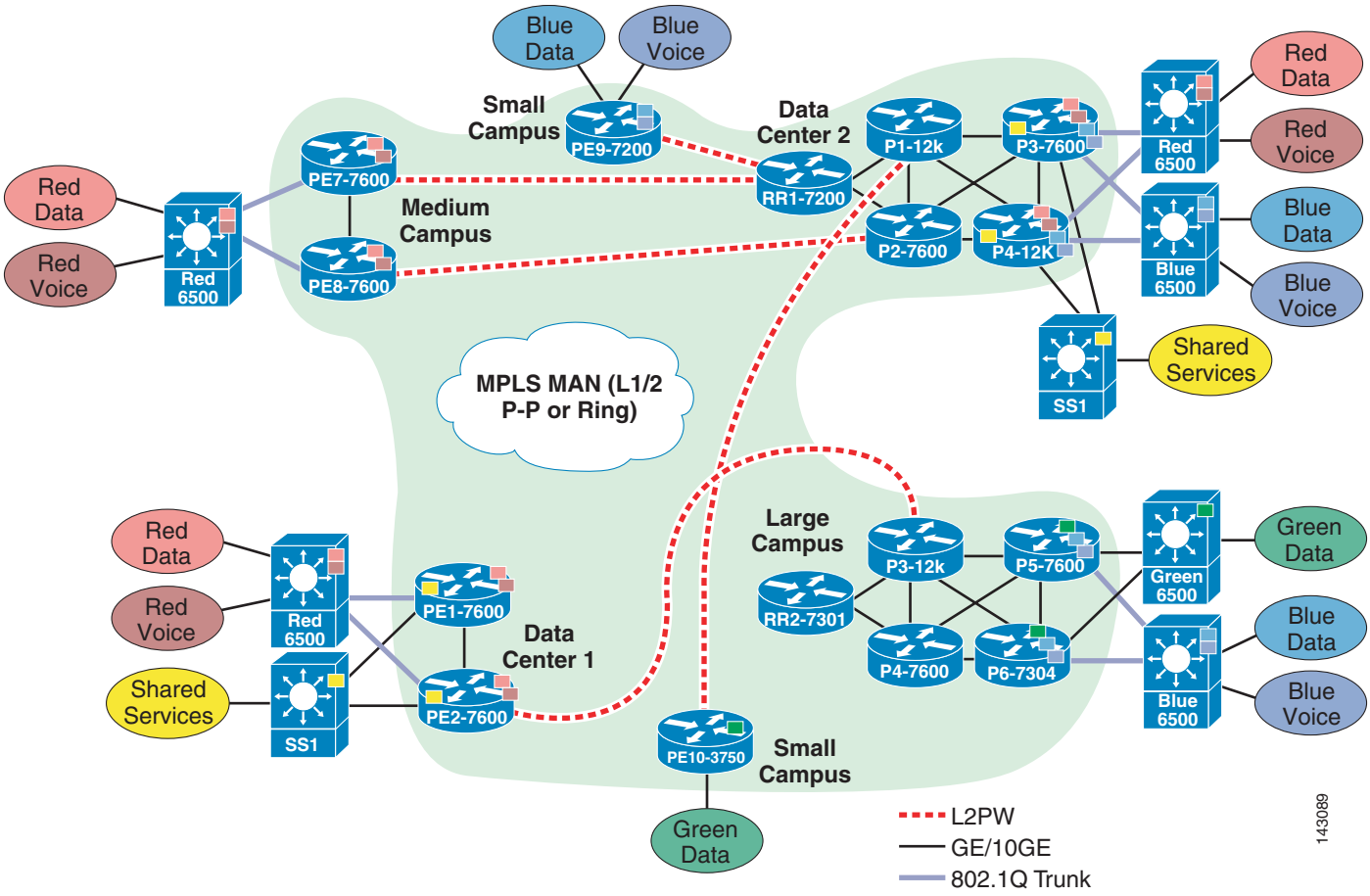
Overview of the Cisco IP Telephony Solution Over the Self-Managed MPLS MAN

The Cisco IP Telephony solution is the leading converged network telephony solution for organizations that want to increase productivity and reduce the costs associated with managing and maintaining separate voice and data networks. The flexibility and sophisticated functionality of the Cisco IP network infrastructure provides the framework that permits rapid deployment of emerging applications such as desktop IP telephony, unified messaging, video telephony, desktop collaboration, enterprise application integration with IP phone displays, and collaborative IP contact centers. These applications enhance productivity and increase enterprise revenues.

The self-managed MPLS MAN enables enterprise customers to begin migration to a more scalable and more efficiently manageable network. It also enables multiple virtual networks on a single network infrastructure (i.e., segmentation).

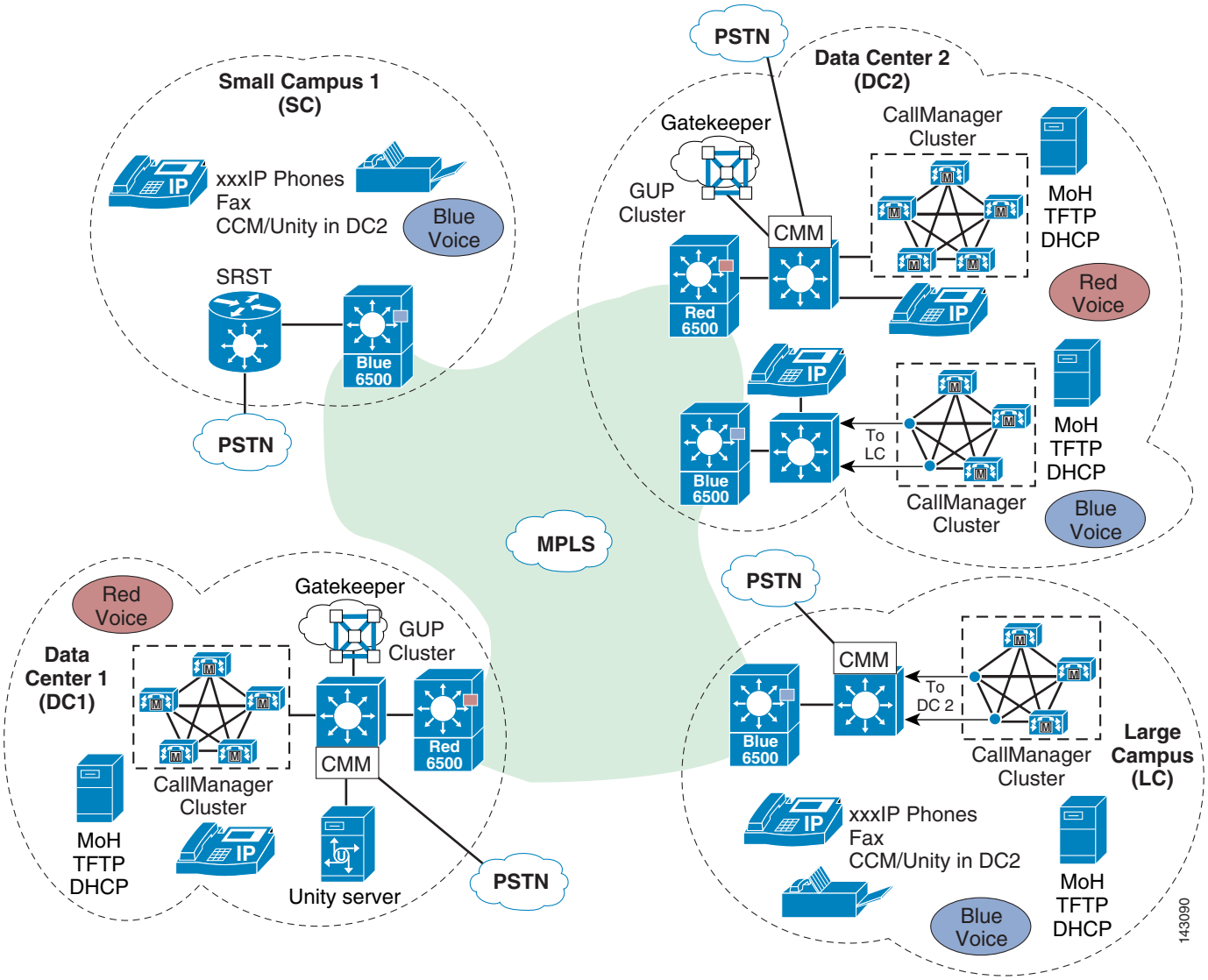
A typical self-managed MPLS MAN allows various organizations and applications to be completely segmented as shown in Figure 6-1.

Figure 6-1 Self-Managed MPLS MAN



This same network with the addition of IP Communications running basic voice servicesis shown in Figure 6-2.

Figure 6-2 Adding IP Communications



In this environment, the network supports multiple organizations segmented across the MPLS MAN segregated by MPLS VPNs. The MPLS VPNs for the voice and data applications are completely segmented with no inter-VPN communications provided.

The foundation architecture of the Cisco IP Telephony solution integrated with the self-managed MPLS MAN includes the following major components:

- Cisco IP network infrastructure
- QoS
- Call processing agent
- Communication endpoints
- Applications

Cisco IP Network Infrastructure

The network infrastructure includes public switched telephone network (PSTN) gateways, analog phone support, and digital signal processor (DSP) farms. The infrastructure can support multiple client types such as hardware phones, software phones, and video devices. The infrastructure also includes the interfaces and features necessary to integrate legacy PBX, voicemail, and directory systems. Typical products used to build the infrastructure include Cisco voice gateways (non-routing, routing, and integrated), Cisco IOS and Cisco Catalyst switches, and Cisco routers.

Quality of Service

Voice, as a class of IP network traffic, has strict requirements concerning packet loss, delay, and delay variation (also known as jitter). To meet these requirements for voice traffic across the MAN, the Cisco IP Telephony solution includes QoS features such as classification, policing, and queuing.

The QoS components of the Cisco IP Telephony solution are provided through the rich IP traffic management, queuing, and policing capabilities of the Cisco IP network infrastructure. Key elements of this infrastructure that enable QoS for IP telephony include:

- Traffic marking
- Enhanced queuing services
- Policing
- Call admission control

Future phases of the design guide will discuss additional QoS tools required to support the WAN and inter-VPN communications.

Call Processing Agent

Cisco CallManager is the core call processing software for the Cisco IP Telephony solution. It builds call processing capabilities on top of the Cisco IP network infrastructure. Cisco CallManager software extends enterprise telephony features and capabilities to packet telephony network devices such as IP phones, media processing devices, Voice over IP (VoIP) gateways, and multimedia applications.

You can deploy the call processing capabilities of Cisco CallManager in the self-managed MPLS MAN according to one of the following models, depending on the size and functional requirements of your enterprise MAN:

- Multi-site MPLS MAN model with distributed call processing—Each site has its own Cisco CallManager cluster for call processing. Communication between sites normally takes place over the IP MPLS MAN, with the PSTN serving as a backup voice path. With this model, you can interconnect any number of sites across the IP MPLS MAN.
- Clustering over the MPLS MAN—You may deploy a single Cisco CallManager cluster across multiple sites that are connected by an MPLS MAN with QoS features enabled. To provide call processing redundancy, you can deploy backup servers either locally at each site or at a remote site across the MPLS MAN. Clustering over the MPLS MAN is well suited as a disaster recovery plan for business continuance sites, perhaps between two data centers.
- Multi-site MPLS MAN model with centralized call processing— The Cisco CallManager cluster resides at a main campus data center. Communication with other offices normally takes place over the MPLS MAN. If either the central site or the MPLS MAN is down, the remote sites can continue to have service through a feature called Survivable Remote Site Telephony (SRST) that runs on Cisco IOS gateways. The remote sites can also place calls over the PSTN if the MPLS MAN is temporarily oversubscribed.

- Hybrid centralized/distributed deployment model across MPLS MAN

The remaining sections of this chapter explain how to apply these deployment models in designing your Cisco IP Telephony network on the self-managed MPLS MAN.

Communication Endpoints

A communication endpoint is a user instrument such as a desk phone or even a software phone application that runs on a PC. In the IP environment, each phone has an Ethernet connection. IP phones have all the functions you expect from a telephone, as well as more advanced features such as the ability to access WWW sites.

In addition to various models of desktop Cisco IP Phones, IP telephony endpoints include the following devices:

- Software-based IP phones

  Cisco IP Communicator and Cisco Softphone are desktop applications that turn your computer into a full-featured IP phone with the added advantages of call tracking, desktop collaboration, and one-click dialing from online directories. Cisco software-based IP phones offer users the great benefit of having a portable office IP phone to use anywhere an Internet connection is available.

- Video telephony endpoints

  Video telephony capability is now fully integrated with Cisco CallManager Release 4.0 and later. In addition, Cisco VT Advantage introduces a Windows-based application and USB camera that can be installed on a Microsoft Windows 2000 or Windows XP personal computer. When the PC is physically connected to the PC port on a Cisco IP Phone 7940, 7960, or 7970, users can make video calls from their IP phones simply by dialing the extension number of another video device on the network.

  Several new third-party video devices are also compatible with the Cisco IP Video Telephony solution.

- Wireless IP Phones

  The Cisco 7920 Wireless IP Phone extends the Cisco family of IP phones from 10/100 Ethernet to 802.11 wireless LAN (WLAN). The Cisco 7920 Wireless IP Phone provides multiple line appearances with functionality similar to existing Cisco 7900 Series IP Phones. In addition, the Cisco 7920 phone provides enhanced WLAN security and Quality of Service (QoS) for operation in 802.11b networks. The Cisco 7920 phone also provides support for XML-based data access and services.

Applications

Voice and video applications build upon the call processing infrastructure to enhance the end-to-end capabilities of the Cisco IP Telephony solution by adding sophisticated telephony and converged network features, such as:

- Unity Messaging

  Cisco Unity delivers powerful unified messaging (email, voice, and fax messages sent to one inbox) and intelligent voice messaging (full-featured voicemail providing advanced functionality) to improve communications, boost productivity, and enhance customer service capabilities across your organization. With Cisco Unity Unified Messaging, you can listen to your email over the phone, check voice messages from the Internet, and (when integrated with a supported third-party fax server) send faxes anywhere.

- Extension mobility

The Cisco CallManager Extension Mobility feature allows users within a Cisco CallManager cluster to configure any Cisco IP Phone 7970, 7960, or 7940 as their own, temporarily, by logging in to that phone. When a user logs in, the phone adopts that user personal phone number(s), speed dials, service links, and other user-specific properties. After logout, the phone reverts to the original user profile. With Cisco CallManager Extension Mobility, several employees can share office space on a rotational basis instead of having a designated office.

- Cisco MeetingPlace

  Cisco MeetingPlace is a complete rich-media conferencing solution that integrates voice, video, and web conferencing capabilities to make remote meetings as natural and effective as face-to-face meetings. In a single step, meeting organizers can schedule voice, video, and web resources through the MeetingPlace web interface, an IP phone, or their Microsoft Outlook or Lotus Notes calendars. Meeting invitees automatically receive notification by email or calendar invitation and can attend rich-media conferences with a single click. With instant messaging applications widely adopted in the workplace, Cisco MeetingPlace also enables users to initiate rich-media conferences easily from common instant messaging clients such as America Online (AOL) Messenger, Lotus Sametime, MSN Messenger, and Yahoo Messenger.

- Web services for Cisco IP Phones

  You can use Cisco IP Phones, such as the Cisco IP Phone 7960 or 7940, to deploy customized client services with which users can interact via the keypad and display. You can create applications for Cisco IP Phone services by using the eXtensible Markup Language (XML) application programming interface (API) and deploy them using the HTTP protocol from standard web servers, such as Microsoft IIS. Some typical services that can be provided through a Cisco IP Phone include a full conferencing interface, the ability to manage data records even if no PC is available, and the ability to display employee alerts, clocks, stock market information, customer contact information, daily schedules, and so forth.

- Cisco IP Contact Center (IPCC) Express

  Cisco IPCC Express is a tightly integrated contact center solution providing three primary functions: interactive voice response (IVR), automatic call distribution (ACD), and computer telephony integration (CTI). The IVR function provides IVR ports to interact with callers by way of either DTMF or speech input. The ACD function provides the ability to intelligently route and queue calls to agents. The CTI function enables call data to be “popped” onto the agent desktop. The IPCC Express software runs on approved Cisco MCS, Hewlett-Packard, or IBM servers and requires interaction with Cisco CallManager.

- Cisco IP Contact Center (IPCC) Enterprise Edition

  Cisco IPCC Enterprise Edition delivers intelligent call routing, network-to-desktop computer telephony integration (CTI), and multi-channel contact management to contact center agents anywhere in the enterprise. The IPCC software profiles each customer using contact-related data such as dialed number and calling line ID, caller-entered digits, data submitted on a Web form, and information obtained from a customer profile database lookup. Simultaneously, the system monitors the resources available in the contact center to meet customer needs, including agent skills and availability, IVR status, queue lengths, and so on. This combination of customer and contact center data is processed through user-defined routing scripts that graphically reflect company business rules, thus enabling Cisco IPCC to route each contact to the optimum resource anywhere in the enterprise.

IP Telephony Deployment Models over the Self-Managed MPLS MAN

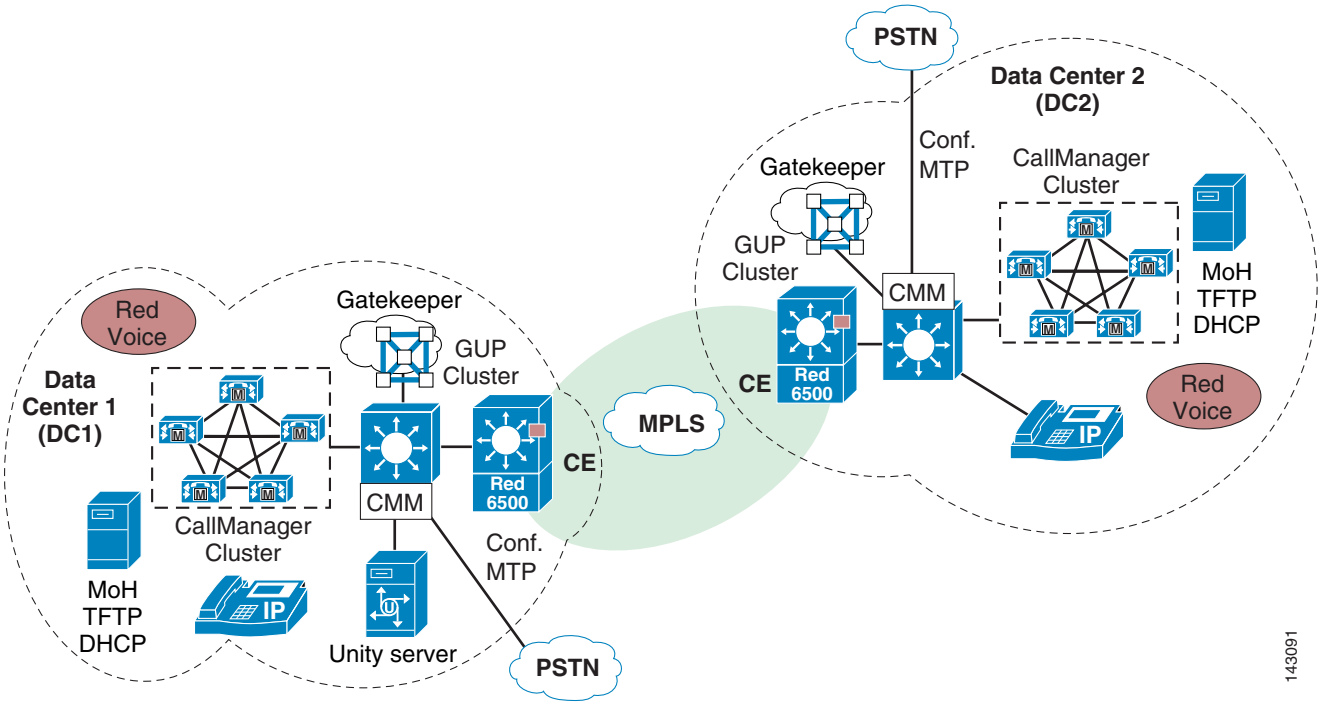
Each Cisco IP Telephony solution is based on one of the following main deployment models described in this chapter:

- Multi-site MPLS MAN model with distributed call processing
- Clustering over the MPLS MAN
- Multi-site MPLS MAN model with centralized call processing
- Hybrid centralized/distributed deployment model across MPLS MAN

Multi-Site MPLS MAN Model with Distributed Call Processing

The multi-site MPLS MAN model with distributed call processing consists of multiple independent sites, each with its own call processing agent connected via the MPLS MAN that carries voice traffic between the distributed sites. Figure 6-3 shows a typical distributed call processing deployment.

Figure 6-3 Typical Distributed Call Processing Deployment



Each site in the distributed call processing model uses its own CallManager cluster for local call processing.

A MPLS MAN interconnects all the distributed call processing sites. Typically, the PSTN is used for off-net calling and serves as a backup connection between the sites in case the MPLS MAN connection fails or does not have any more available bandwidth.

Benefits of the Distributed Call Processing Model

The multi-site MPLS MAN model with distributed call processing provides these benefits:

- Use of the MPLS network for call routing

- Maximum utilization of available bandwidth by allowing voice traffic to share the IP WAN with other types of traffic
- No loss of functionality during a MPLS failure because there is a call processing agent at each site
- Scalability to hundreds of sites

Best Practices for the Distributed Call Processing Model

Follow these guidelines and best practices when implementing the multi-site Distributed Call Processing model:

- Provide a highly available, fault-tolerant infrastructure based on a common infrastructure philosophy. A sound infrastructure is essential for easier migration to IP telephony, integration with applications such as video streaming, and video conferencing.
- Use G.711 codecs for all endpoints. This practice eliminates the consumption of digital signal processor (DSP) resources for transcoding so those resources can be allocated to other functions such as conferencing and Media Termination Points (MTPs).
- Use Media Gateway Control Protocol (MGCP) gateways for the PSTN if you do *not* require H.323 functionality. This practice simplifies the dial plan configuration. H.323 might be required to support specific functionality not offered with MGCP, such as support for Signaling System 7 (SS7) or Non-Facility Associated Signaling (NFAS).
- Implement the recommended network infrastructure for high availability, connectivity options for phones (in-line power), Quality of Service (QoS) mechanisms, and security. See the Enterprise IP Telephony SRND at the following URL:

  http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html
- Follow the provisioning recommendations listed in the Enterprise IP telephony SRND chapter on Call Processing:
  http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html

Gatekeeper or Session Initiation Protocol (SIP) proxy servers are among the key elements in the multi-site MPLS MAN model with distributed call processing. They each provide dial plan resolution, with the gatekeeper also providing call admission control, although bandwidth is not normally a limitation for transporting voice in a MAN design. A gatekeeper is an H.323 device that provides call admission control and E.164 dial plan resolution.

In multi-site deployments where a Cisco CallManager cluster is present at each site and the sites are linked through the MPLS MAN, a gatekeeper can provide call admission control between the sites, with each site being placed in a different gatekeeper zone.

Bandwidth would be a rare instance because bandwidth is not normally an issue when interconnecting telephony sites across the MPLS MAN; however, there is such a requirement for Call Admission Control, when all the available bandwidth that is provisioned for voice between particular sites has been utilized, you can provide automatic failover to the PSTN using the route list and route group construct for the route patterns that connect each cluster to the gatekeeper. For more detail on automatic failover see the Enterprise IP Telephony SRND at the following URL:
http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

The following best practices apply to the use of a gatekeeper:

- Use a Cisco IOS gatekeeper to provide call admission control into and out of each site.

- To provide high availability of the gatekeeper, use Hot Standby Router Protocol (HSRP) gatekeeper pairs, gatekeeper clustering, and alternate gatekeeper support. In addition, use multiple gatekeepers to provide redundancy within the network.
- Size the platforms appropriately to ensure that performance and capacity requirements can be met.
- Because this is a MAN deployment and bandwidth is plentiful, use the single G.711 codec on the MPLS MAN.
- Gatekeeper networks can scale to hundreds of sites and the design is limited only by the MAN topology.

For more information on the various functions performed by gatekeepers, see the following sections such as scalability, redundancy, and dial plan resolution in the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

SIP devices provide resolution of E.164 numbers as well as SIP uniform resource identifiers (URIs) to enable endpoints to place calls to each other. Cisco CallManager supports the use of E.164 numbers only.

The following best practices apply to the use of SIP proxies:

- Provide adequate redundancy for the SIP proxies.
- Ensure that the SIP proxies have the capacity for the call rate and number of calls required in the network.

Note Planning for call admission control is outside the scope of this document.

Clustering over the MPLS MAN

You may deploy a single Cisco CallManager cluster across multiple sites, such as two data centers, that are connected by the MPLS MAN. This section provides a brief overview of clustering over the MAN.

For further information, see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Clustering over the WAN can support two types of deployments:

- Local Failover Deployment Model

  Local failover requires that you place the Cisco CallManager subscriber and backup servers at the same site with no WAN between them. This deployment model is ideal for two to four sites with Cisco CallManager

- Remote Failover Deployment Model

  Remote failover allows you to deploy the backup servers across the MPLS MAN. Using this deployment model, you may have up to eight sites with Cisco CallManager subscribers being backed up by Cisco CallManager subscribers at another site.

You can also use a combination of the two deployment models to satisfy specific site requirements. For example, two main sites may each have primary and backup subscribers, with another two sites containing only a primary server each and using either shared backups or dedicated backups at the two main sites.

The key advantages of clustering over the WAN are:

- Single point of administration for users for all sites within the cluster

- Feature transparency
- Shared line appearances
- Extension mobility within the cluster
- Unified dial plan

These features make this solution ideal as a disaster recovery plan for business continuance sites or as a single solution for up to eight small or medium sites.

MPLS MAN Considerations

For clustering over the MPLS MAN to be successful, you must carefully plan, design, and implement various characteristics of the MAN itself. The Intra-Cluster Communication Signaling (ICCS) between Cisco CallManager servers consists of many traffic types. The ICCS traffic types are classified as either Priority or Best Effort. Priority ICCS traffic is marked with IP Precedence 3 (DSCP 26 or PHB AF31 in Cisco CallManager releases before 4.0 and DSCP 24 or PHB CS3 for Release 4.0 and later). Best Effort ICCS traffic is marked with IP Precedence 0 (DSCP 0 or PHB BE). The various types of ICCS traffic are described in the subsequent Intra-Cluster Communications section, which also provides further guidelines for provisioning.

Delay

Because delay is minimal across the MPLS MAN, details are not presented in this section. However, it is still important to understand the delay requirements for designing a cluster across the MAN/WAN. For details, see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Jitter

Jitter is the varying delay that packets incur through the network because of processing, queue, buffer, congestion, or path variation delay. Jitter for the IP Precedence 3 ICCS traffic must be minimized using QoS features.

Packet Loss and Errors

The network should be engineered for zero percent packet loss and errors for all ICCS, especially the priority ICCS traffic, because packet loss and errors have adverse effects on the real-time call processing within the cluster.

Bandwidth

Bandwidth is not normally a cause of concern in the MPLS MAN environment. Typically, you provision the correct amount of bandwidth between each server for the expected call volume, type of devices, and number of devices. This bandwidth is in addition to any other bandwidth for other applications sharing the network, including voice and video traffic between the sites. The bandwidth provisioned must have QoS enabled to provide the prioritization and scheduling for the different classes of traffic. The general rule of thumb for bandwidth is to over-provision and under-subscribe.

Quality of Service

The network infrastructure relies on QoS engineering to provide consistent and predictable end-to-end levels of service for traffic. Neither QoS nor bandwidth alone is the solution; rather, QoS-enabled bandwidth must be engineered into the network infrastructure.

Intra-Cluster Communications

In general, intra-cluster communications means all traffic between servers. There is also a real-time protocol called Intra-Cluster Communication Signaling (ICCS) that provides the communications with the Cisco CallManager Service process that is at the heart of the call processing in each server or node within the cluster.

The intra-cluster traffic between the servers consists of:

- Database traffic from the SQL database that provides the main configuration information. The SQL database is replicated from the publisher server to all other servers in the cluster using Best Effort. The SQL traffic may be re-prioritized in line with Cisco QoS recommendations to a higher priority data service (for example, IP Precedence 1 if required by the particular business needs). An example of this is extensive use of Extension Mobility that relies on SQL database configuration.
- Directory traffic from the Lightweight Directory Access Protocol (LDAP) directory provides user and application authentication and some additional specific user or application configuration information. LDAP traffic is sent Best Effort by default.
- ICCS real-time traffic, which consists of signaling, call admission control, and other information regarding calls as they are initiated and completed. ICCS uses a TCP connection between all servers that have the Cisco CallManager Service enabled. The connections are a full mesh between these servers. Because only eight servers may have the Cisco CallManager Service enabled in a cluster, there may be up to seven connections on each server. This traffic is priority ICCS traffic and is marked dependant on release and service parameter configuration.
- CTI Manager real-time traffic is used for CTI devices involved in calls or for controlling or monitoring other third-party devices on the Cisco CallManager servers. This traffic is marked as priority ICCS traffic and exists between the Cisco CallManager server with the CTI Manager and the Cisco CallManager server with the CTI device.

Failover between Subscriber Servers

With Cisco CallManager Release 3.1 and 3.2, failover behavior is dependant on the reachability of the publisher and the delay between the subscriber and the publisher. If the publisher is reachable, the subscriber requests the relevant device configuration records directly from the publisher during device registration. The round-trip delay and the available bandwidth for the SQL database traffic affects the speed of registrations. The effect of this is that failover for devices at remote locations to the publisher may experience delays of approximately 20 minutes before all devices on a full server complete the failover process. If the publisher is unreachable during failover, the subscriber uses its own most recent copy of the database for the configuration information. Because there is no incurred delay for the subscriber to access its own database, the failover time in this case is approximately five minutes for a full server.

With Cisco CallManager Release 3.3 and higher, the impact of delay to the publisher is minimized during the failover period because the configuration information is cached during initialization or boot-up time. The effect is that the Cisco CallManagers might take longer to start up initially; however any subsequent failover and failback is not affected by the delay in accessing the publisher database.

Cisco CallManager Publisher

The publisher replicates a read-only copy of the master database to all other servers in the cluster. If changes are made in the publisher master database during a period when another server in the cluster is unreachable, the publisher replicates the updated database when communications are re-established.

During any period when the publisher is unreachable or offline, no changes can be made to the configuration database. All subscriber databases are read-only and may not be modified. Most normal operations of the cluster are not affected during this period, including:

- Call processing
- Failover
- Installation registration of previously configured devices

There are some features and functions that require access to the master database on the publisher because they make modifications to records and therefore need write access. The publisher is the only server in a Cisco CallManager cluster that has a read and write configuration database. The main features and functions that require access to the publisher for write access include:

- Configuration additions, changes, and deletions
- Extension Mobility
- User speed dials
- Cisco CallManager User page options requiring the database
- Cisco CallManager software upgrades
- Call Forward All changes
- Message Waiting Indicator (MWI) state

Other services or applications might also be affected and their ability to function without the publisher should be verified when deployed.

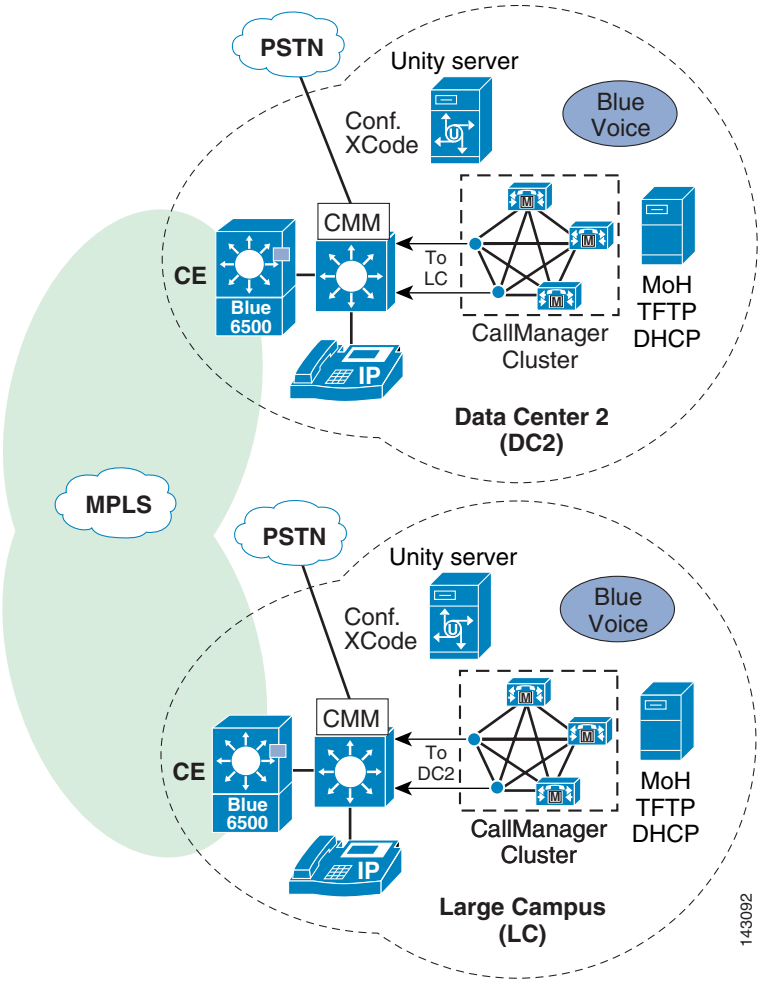
Call Detail Records (CDR)

Call detail records, when enabled, are collected by each subscriber and uploaded to the publisher periodically. During a period when the publisher is unreachable, the CDRs are stored on the subscriber local hard disk. When connectivity is re-established to the publisher, all outstanding CDRs are uploaded to the publisher.

Local Failover Deployment Model

The local failover deployment model provides the most resilience for clustering the sites in this model contains at least one primary Cisco CallManager subscriber and one backup subscriber. This configuration can support up to four sites. The maximum number devices is dependant on the quantity and type of servers deployed. The maximum IP phones for all sites is 30,000 (see Figure 6-4).

Figure 6-4 Local Failover Deployment Model

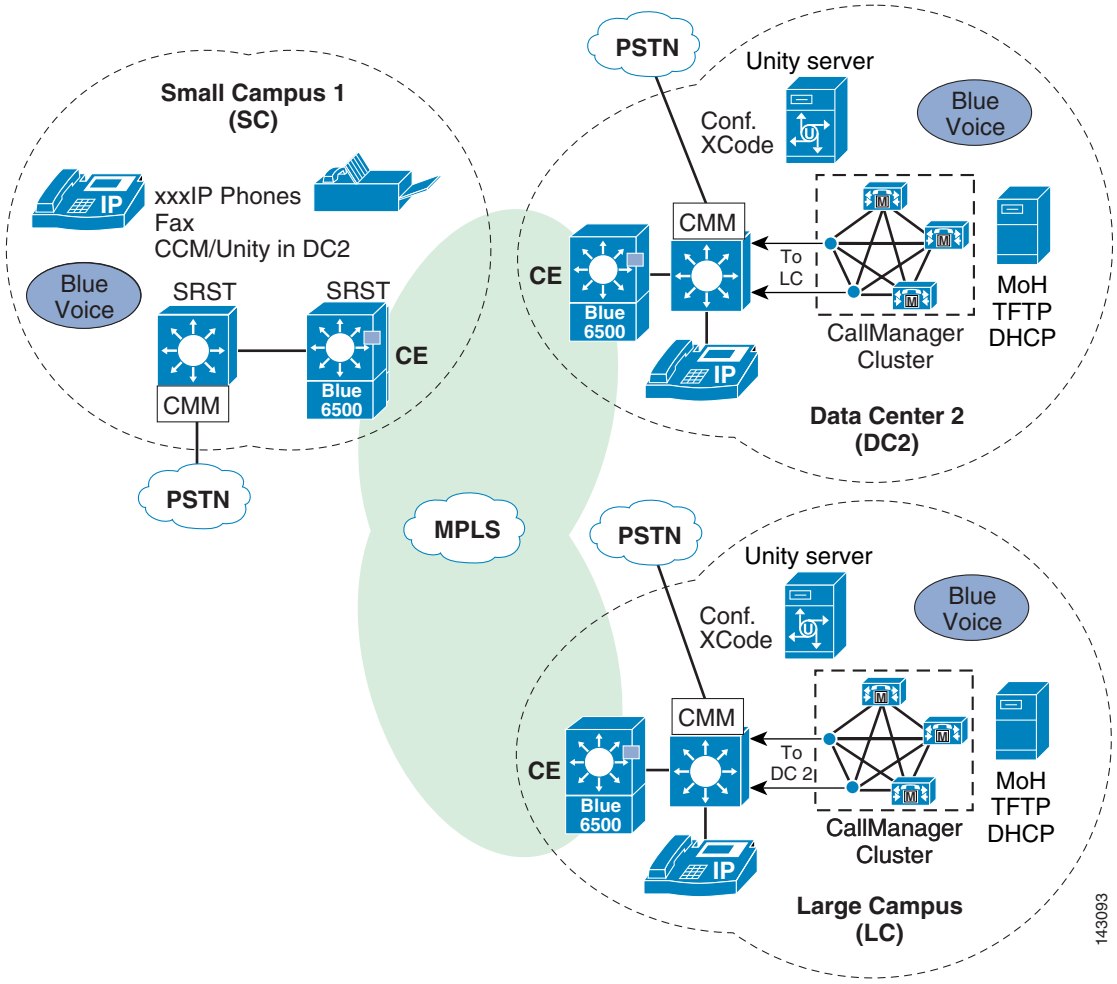


Observe the following guidelines when implementing the local failover model:

- Configure each site to contain at least one primary Cisco CallManager subscriber and one backup subscriber.
- Configure Cisco CallManager *groups* and *device pools* to allow devices within the site to register with only the servers at that site under all conditions.
- Cisco highly recommends that you replicate key services (TFTP, DNS, DHCP, LDAP, and IP Phone services), all media resources (conference bridges and music on hold), and gateways at each site to provide the highest level of resiliency. You can also extend this practice to include a voicemail system at each site.
- Under a failure condition, sites without access to the publisher database lose some functionality:
  - System administration at the local site is not able to add, modify, or delete any part of the configuration.
  - Extension mobility users are not able to log in or log out of the IP phones.
  - Changes to Call Forward All are not allowed.

- Under MPLS MAN failure conditions, calls made to phone numbers that are not currently communicating with the subscriber placing the call result in either a fast-busy tone or a call forward (possibly to voicemail, depending on the location of the phone number to which they are being forwarded). During this condition, users should manually dial those numbers via the PSTN.
- Every 10,000 busy hour call attempts (BHCA) between sites that are clustered over the WAN requires 900 kbps of bandwidth for ICCS. This is a minimum bandwidth requirement and bandwidth is allocated in multiples of 900 kbps. The ICCS traffic types are classified as either Priority or Best Effort. Priority ICCS traffic is marked with IP Precedence 3 (DSCP 26 or PHB AF31 in Cisco CallManager releases before 4.0 and DSCP 24 or PHB CS3 for Release 4.0 and later). Best Effort ICCS traffic is marked with IP Precedence 0 (DSCP 0 or PHB BE).
- The local failover model requires Cisco CallManager Release 3.1 or later.
- Because the MAN is based on MPLS, any site connected to the MAN that does not have a local CallManager subscriber for call processing may have those telephony devices register with any CallManager subscriber at either of the main CallManager cluster sites (see Figure 6-5).

Figure 6-5 Remote Subscription to CallManager



- During a software upgrade, all servers in the cluster should be upgraded during the same maintenance period using the standard upgrade procedures outlined in the software release notes.

Cisco CallManager Provisioning for Local Failover

Provisioning of the Cisco CallManager cluster for the local failover model should follow the design guidelines for capacities outlined in the chapter on Call Processing in the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

If voice or video calls are allowed across the MPLS MAN between the sites, then you must configure Cisco CallManager *locations*, in addition to the default location for the other sites, to provide call admission control between the sites. Even though the bandwidth is more than likely over-provisioned across the MPLS MAN for the number of devices, it is still best practice to configure call admission control based on locations. If the locations-based call admission control rejects a call, automatic failover to the PSTN can be provided by the automated alternate routing (AAR) feature.

To improve redundancy and upgrade times, Cisco recommends that you enable the Cisco TFTP service on at least one of the Cisco CallManager servers at each location. You can run the TFTP service on either a publisher or a subscriber server, depending on the site and the available capacity of the server. The TFTP server option must be correctly set on the DHCP servers for each site. If DHCP is not in use or the TFTP server is manually configured, you should configure the correct address for the site.

Other services which may affect normal operation of Cisco CallManager during MPLS MAN outages should also be replicated at all sites to ensure uninterrupted service. These services include DHCP servers, DNS servers, corporate directories, and IP phone services. On each DHCP server, set the DNS server address correctly for each location.

Gateways for Local Failover

Normally, gateways should be provided at all sites for access to the PSTN. The device pools should be configured to register the gateways with the Cisco CallManager servers at the same site. Partitions and calling search spaces should also be configured to select the local gateways at the site as the first choice for PSTN access and the other site gateways as a second choice for overflow. Take special care to ensure emergency service access at each site.

You can centralize access to the PSTN gateways if access is not required during a MAN failure. For E911 requirements, additional gateways might be needed at each site.

Voicemail for Local Failover

Cisco Unity or other voicemail systems can be deployed at all sites and integrated into the Cisco CallManager cluster. This configuration provides voicemail access even during a MAN failure and without using the PSTN.

Using Voice Mail Profiles, you can allocate the correct voicemail system for the site to the IP phones in the same location. You can configure a maximum of four voicemail systems per cluster that use the SMDI protocol, which are attached directly to the COM port on a subscriber and that use the Cisco Messaging Interface (CMI).

Music on Hold and Media Resources for Local Failover

Music on hold (MoH) servers and other media resources such as conference bridges should be provisioned at each site with sufficient capacity for the type and number of users. Through the use of media resource groups (MRGs) and media resource group lists (MRGLs), media resources are provided by the on-site resource and are available during a MAN failure.

The remote failover deployment model provides flexibility for the placement of backup servers. Each of the sites contains at least one primary Cisco CallManager subscriber and may or may not have a backup subscriber. This model allows for a deployment of up to eight sites with IP phones and other devices normally registered to a local subscriber when using 1:1 redundancy and the 50/50 load balancing option described in the Enterprise IP Telephone SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Backup subscribers are located across the MPLS MAN at one or more of the other sites.

When implementing the remote failover model, observe all guidelines for the local failover model with the following modifications:

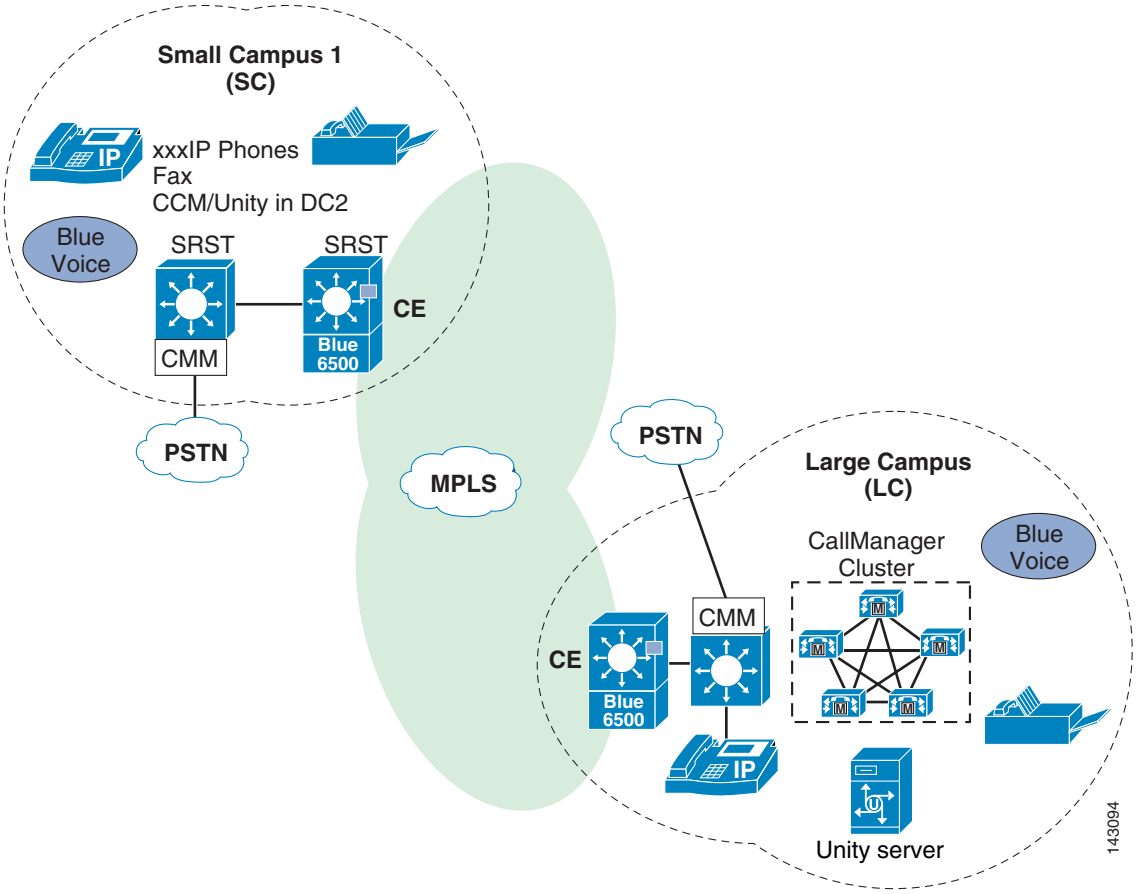
- Configure each site to contain at least one primary Cisco CallManager subscriber and an optional backup subscriber as desired.
- You may configure Cisco CallManager *groups* and *device pools* to allow devices to register with servers over the MPLS MAN.

Multi-Site MPLS MAN Model with Centralized Call Processing

The multi-site MPLS MAN model with centralized call processing consists of a single call processing agent that provides services for many sites and uses the MPLS MAN to transport IP telephony traffic between the sites. The MPLS MAN also carries call control signaling between the central site and the remote sites.

Figure 6-6 illustrates a typical centralized call processing deployment, with a Cisco CallManager cluster as the call processing agent at a data center central site and an IP MPLS MAN with QoS enabled to connect one or multiple additional sites.

Figure 6-6 Multi-site MPLS MAN Model with Centralized Call Processing



The campus sites remote from CallManager rely on the centralized Cisco CallManager cluster to handle their call processing. Applications such as voicemail are typically centralized as well to reduce the overall costs of administration and maintenance.

Note In each solution for the centralized call processing model presented in this document, the various sites connect to a MPLS MAN with QoS configured.

Routers and switches that reside at the edges require QoS mechanisms, such as priority queuing and policing to protect the voice traffic from the data traffic across the MPLS MAN. In addition, a call admission control scheme is needed to avoid oversubscribing the links with voice traffic and deteriorating the quality of established calls.

For centralized call processing deployments, the *locations* construct within Cisco CallManager provides call admission control. See the section on Cisco CallManager Locations; for more information on locations see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

A variety of Cisco gateways can provide the remote sites with PSTN access. If all the available bandwidth allocated for voice on the MPLS MAN has been consumed, the system uses AAR to re-route calls between sites across the PSTN.

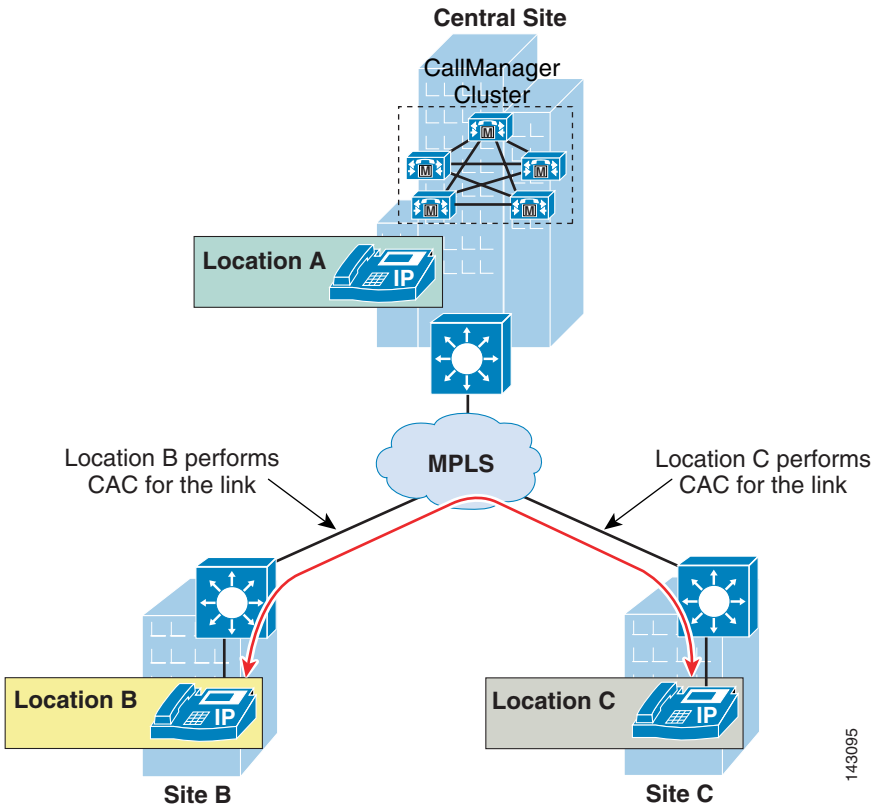
The Survivable Remote Site Telephony (SRST) feature, available on Cisco IOS gateways, provides call processing at the remote offices from CallManager in the event of a MPLS MAN failure. Users at each remote site can dial the PSTN access code and place their calls through the PSTN.

Best Practices for the Centralized Call Processing Model

Follow these guidelines and best practices when implementing the multi-site MPLS MAN model with centralized call processing:

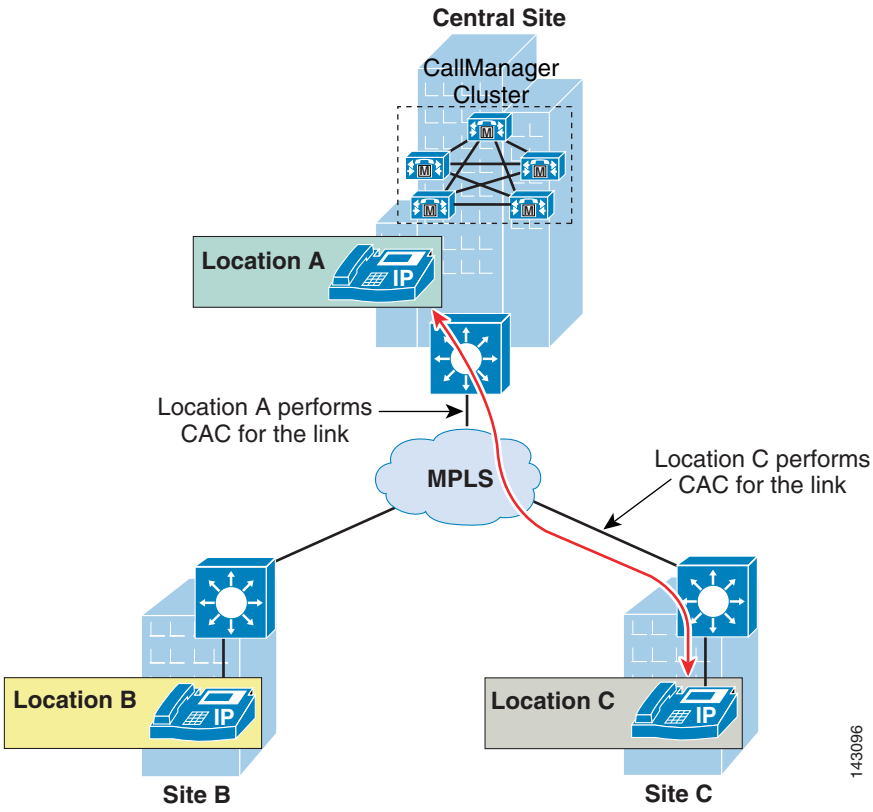
- In single-cluster centralized call processing deployments, the call admission control (CAC) function is performed by the *locations* construct within Cisco CallManager. Although in a MPLS MAN environment there would most likely not be bandwidth issues requiring CAC if there is such a requirement follow these best practices.
- With an MPLS network, all sites connected across the MAN are deemed to be adjacent at Layer 3, thus they do not have to rely on the central site for connectivity.

Figure 6-7 CAC for Calls Between Remote Sites



- Also, in an MPLS MAN, the link connecting the central Call Processing site does not aggregate every remote site link. Place all the central site devices in their own call admission control location (that is, not in the <None> location); this configuration requires that call admission control be performed on the central site link independently of the remote site links.

Figure 6-8 CAC for Calls Between Central Site and Remote Sites



- When all the available bandwidth reserved for voice for a particular site has been utilized, you can provide automatic failover to the PSTN using the automated alternate routing (AAR) feature within Cisco CallManager. For more information on AAR, see the section on Automated Alternate Routing in the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.
- Use the locations mechanism in Cisco CallManager to provide call admission control into and out of sites that do not have CallManager servers.
- The locations mechanism works across multiple servers in Cisco CallManager Release 3.1 and later.
- This configuration can support a maximum of 30,000 IP phones when Cisco CallManager runs on the largest supported server.
- The number of IP phones and line appearances supported in SRST mode at each remote site depends on the router or Catalyst gateway module used, the amount of memory installed, and the Cisco IOS release. (For the latest SRST platform and code specifications, see the SRST documentation at http://www.cisco.com.) Generally speaking, however, the choice of whether to adopt a centralized call processing or distributed call processing approach for a given site depends on a number of factors such as:
  - Criticality of the voice network
  - Feature set needs
  - Scalability
  - Ease of management

- Cost

If a distributed call processing model is deemed more suitable for customer business needs, you would include in the design the installation of a local Cisco CallManager cluster at each location or design a CallManager cluster across multiple locations as described above.

Survivable Remote Site Telephony

When deploying IP telephony across a self-managed MPLS MAN with the centralized call processing model, the MAN is highly available based on the robustness to the MPLS design.

However, should there be a catastrophic outage of the network isolating the remote locations from the centralized CallManager cluster, SRST provides high availability for voice services only by providing a subset of the call processing capabilities within the remote office location and enhancing the IP phones with the ability to "re-home" to the call processing functions in the local network infrastructure device if a the MPLS network failure is detected.

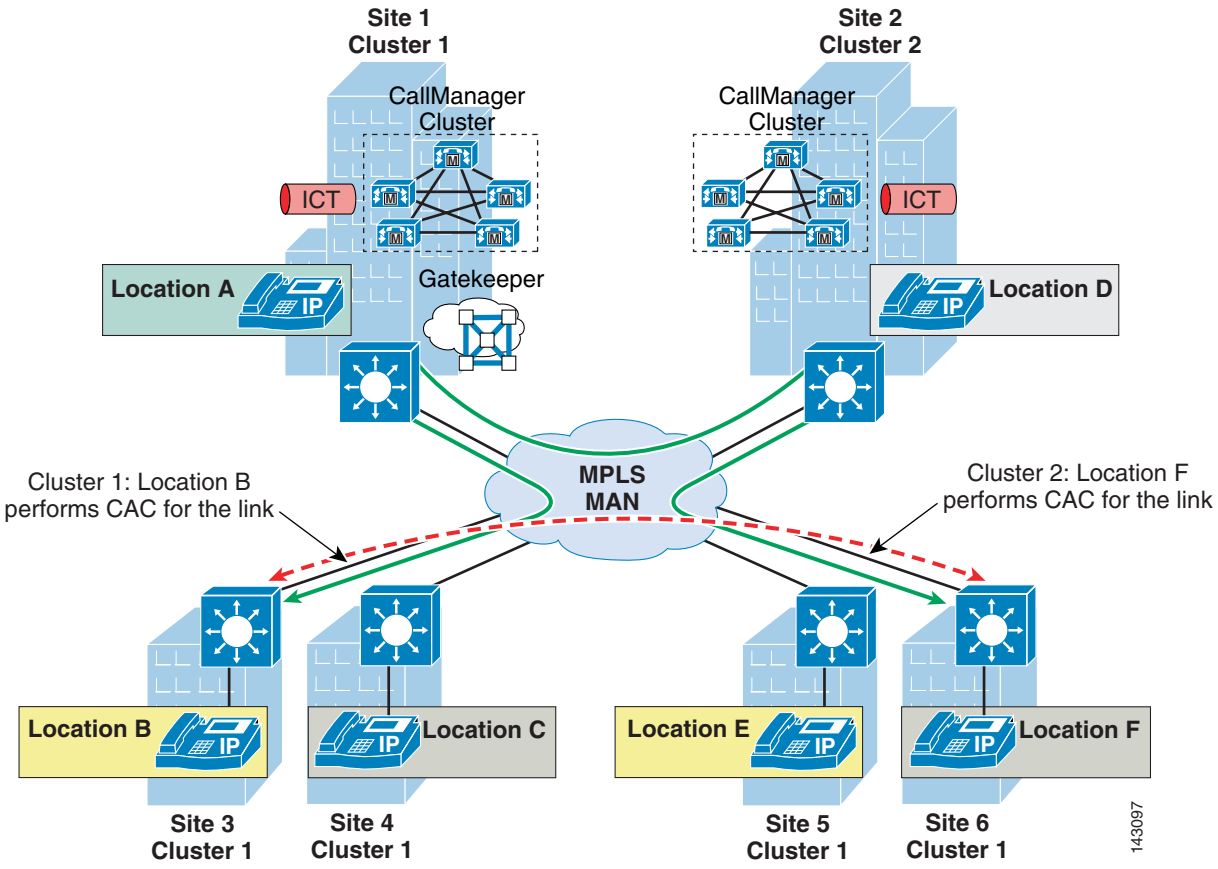
For more detail of SRST functionality in a IP Communications centralized call processing environment, see the following URL:
http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Hybrid Centralized/Distributed Deployment Model across MPLS MAN

For multi-site deployments that combine both centralized and distributed call processing deployment models, a MPLS MAN presents a new situation for inter-cluster calls.

When calls occur between two sites belonging to different clusters, the audio path is established between the two sites directly with no media transiting through each cluster central site. Therefore call admission control is required only on the link at the two remote sites (see Figure 6-9) .

Figure 6-9 CAC for Distributed Deployment Model



As in the purely centralized deployments, devices that terminate media at each site (including the central sites for each cluster) must be placed in an appropriately configured location.

Note that the inter-cluster trunks are purely signaling devices and there is no media transiting through them. Therefore all inter-cluster trunks must be left in location <None>.

In these deployments, a gatekeeper can be used for dial plan resolution between clusters, but a gatekeeper is not recommended for call admission control.

Although bandwidth is normally not an issue to carry voice in a MPLS MAN environment, if the available bandwidth provisioned for voice for a particular site has been used, you can provide automatic failover to the PSTN by using a combination of the following two methods:

- The route list and route group construct for calls across multiple Cisco CallManager clusters
- The automated alternate routing (AAR) feature for calls within a Cisco CallManager more information on AAR. See the “Automated Alternate Routing” section in the Enterprise IP telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Network Infrastructure

This section describes the requirements of the network infrastructure needed to build an IP telephony system in an enterprise environment. Figure 6-10 shows the roles of the various devices that form the network infrastructure.

Figure 6-10 Network Infrastructure

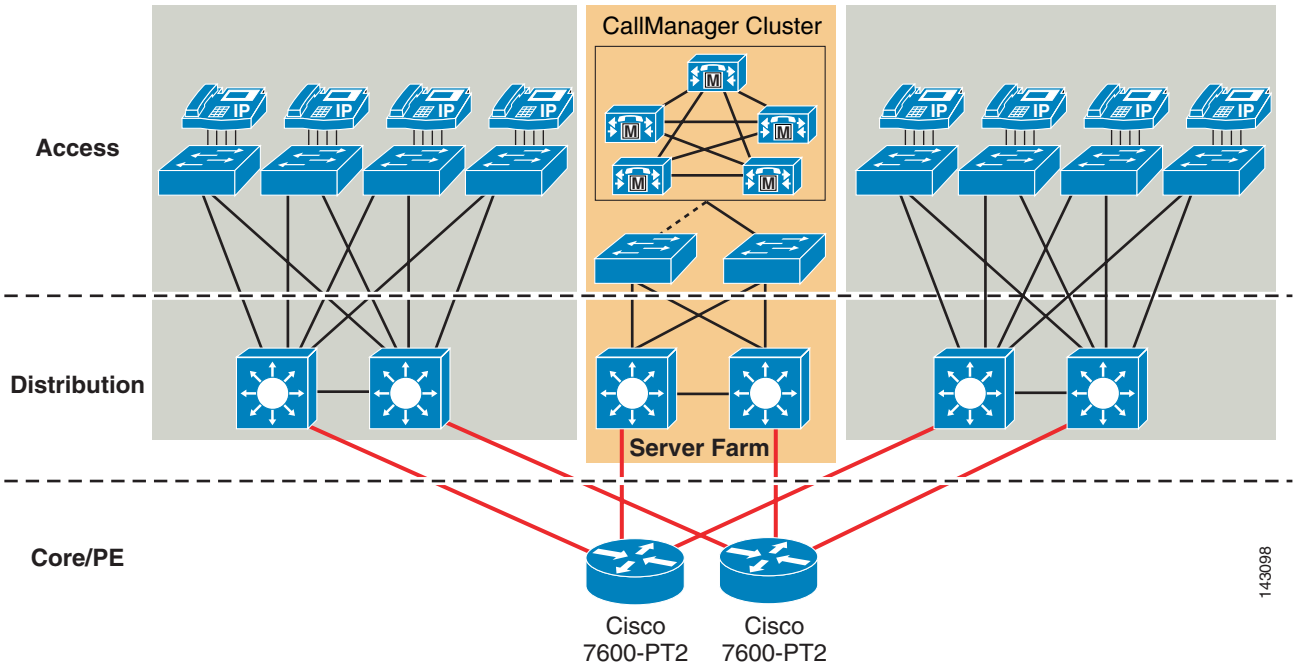


Table 6-1 summarizes the features required to support each of these roles.

Table 6-1 Required Features

Infrastructure Role	Required Feature
Campus access switch	In-line power
	Multiple queue support
	802.1p and 801.2Q
	Fast link convergence
Campus distribution	Multiple queue support
	VRF lite
	Traffic classification
	Traffic re-classification
PE	Multiple queue support
	VRF
	Traffic classification
	Traffic re-classification

Campus Access Layer

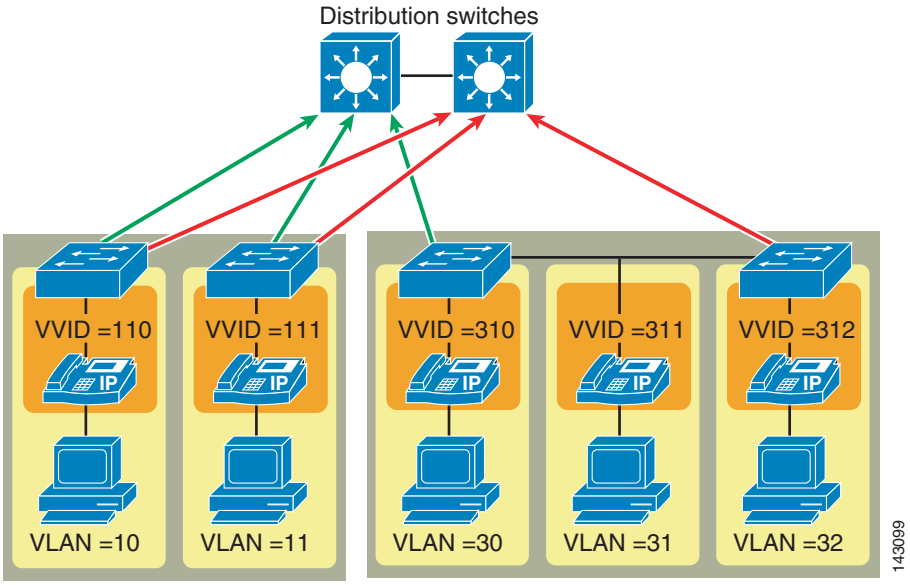
This section focuses on the campus access layer. The remaining blocks and component of self-managed MPLS infrastructure are addressed in the Next Generation MPLS Architecture and QoS sections of this design guide in addition to the appropriate Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

The access layer of the Campus LAN includes the portion of the network from the desktop port(s) to the wiring closet switch.

It is currently required to design the access layer using traditional design methodology to the distribution layer switches because many of the features required for voice applications, for example SRST, Gateways, DSP Resources, and so on, currently are not VRF-aware.

Proper access layer design starts with assigning a single IP subnet per virtual LAN (VLAN). Typically, a VLAN should not span multiple wiring closet switches; that is, a VLAN should have presence in one and only one access layer switch (see Figure 6-11). This practice eliminates topological loops at Layer 2, thus avoiding temporary flow interruptions because of Spanning Tree convergence. However with the introduction of standards-based IEEE 802.1w Rapid Spanning Tree Protocol (RSTP) and 802.1s Multiple Instance Spanning Tree Protocol (MISTP), Spanning Tree can converge at much higher rates.

Figure 6-11 Campus Access Layer



In situations where RSTP and/or MISTP can and have been configured on the access layer switch, there is no need for concern about topological loops. More importantly, confining a VLAN to a single access layer switch also serves to limit the size of the broadcast domain. There is the potential for large numbers of devices within a single VLAN or broadcast domain to generate large amounts of broadcast traffic periodically, which can be problematic. A good rule of thumb is to limit the number of devices per VLAN to about 512, which is equivalent to two Class C subnets (that is, a 23-bit subnet masked Class C address).

When you deploy voice, Cisco recommends that you enable two VLANs at the access layer: a native VLAN for data traffic (VLANs 10, 11, 30, 31, and 32 in Figure 6-11) and a voice VLAN under Cisco IOS or auxiliary VLAN under CatOS for voice traffic (represented by VVIDs 110, 111, 310, 311, and 312 in Figure 6-11).

Separate voice and data VLANs are recommended for the following reasons:

- Address space conservation and voice device protection from external networks
- Private addressing of phones on the voice or auxiliary VLAN ensures address conservation and ensures that phones are not accessible directly via public networks. PCs and servers are typically addressed with publicly-routed subnet addresses; however voice endpoints should be addressed using RFC 1918 private subnet addresses.
- QoS trust boundary extension to voice devices
- QoS trust boundaries can be extended to voice devices without extending these trust boundaries and, in turn, QoS features to PCs and other data devices
- Protection from malicious network attacks
- VLAN access control, 802.1Q, and 802.1p tagging can provide protection for voice devices from malicious internal and external network attacks such as worms, DoS attacks, and attempts by data devices to gain access to priority queues via packet tagging.
- Ease of management and configuration

  Separate VLANs for voice and data devices at the access layer provide ease of management and simplified QoS configuration.

To provide high-quality voice and to take advantage of the full voice feature set, access layer switches should provide support for:

- 802.1Q trunking and 802.1p for proper treatment of Layer 2 CoS packet marking on ports with phones connected
- Multiple egress queues to provide priority queuing of RTP voice packet streams
- The ability to classify or reclassify traffic and establish a network trust boundary
- Inline power capability (although inline power capability is not mandatory, it is highly recommended for the access layer switches)
- Layer 3 awareness and the ability to implement QoS access control lists (these features are required if you are using certain IP telephony endpoints, such as a PC running a softphone application that cannot benefit from an extended trust boundary)

Spanning Tree Protocol (STP)

To minimize convergence times and maximize fault tolerance at Layer 2, enable the following STP features:

- PortFast

  Enable PortFast on all access ports. The phones, PCs, or servers connected to these ports do not forward bridge protocol data units (BPDUs) that could affect STP operation. PortFast ensures that the phone or PC when connected to the port is able to begin receiving and transmitting traffic immediately without having to wait for STP to converge.
- Root Guard or BPDU Guard

  Enable root guard or BPDU guard on all access ports to prevent the introduction of a rogue switch that might attempt to become the Spanning Tree root, thereby causing STP re-convergence events and potentially interrupting network traffic flows. Ports that are set to errdisable state by BPDU guard must either be re-enabled manually or the switch must be configured to re-enable ports automatically from the errdisable state after a configured period of time.
- UplinkFast and BackboneFast

Enable these features where appropriate to ensure that when changes occur on the Layer 2 network, STP converges as rapidly as possible to provide high availability. When using stackable switches such as the Catalyst 2950, 3550, or 3750, enable Cross-Stack UplinkFast (CSUF) to provide fast failover and convergence if a switch in the stack fails.

- UniDirectional Link Detection (UDLD)

  Enable this feature to reduce convergence and downtime on the network when link failures or misbehaviors occur, thus ensuring minimal interruption of network service. UDLD detects and takes out of service links where traffic is flowing in only one direction. This feature prevents defective links from being mistakenly considered as part of the network topology by the Spanning Tree and routing protocols.

Note With the introduction of RSTP 802.1w, features such as PortFast and UplinkFast are not required because these mechanisms are built into this standard. If RSTP has been enabled on the Catalyst switch, these commands are not necessary.

CallManager Server Farm

Cisco CallManager cluster servers, including media resource servers, typically reside in a data center or server farm environment. In addition, centralized gateways and centralized hardware media resources such as conference bridges, DSP or transcoder farms, and media termination points are located in the data center or server farm.

The server farm is typically implemented at the access layer, which must currently be designed using traditional design methodology to the distribution layer switches because many of the features required for voice applications, such as SRST, gateways, DSP resources, and so on are not currently VRF-aware.

Because these servers and resources are critical to voice networks, Cisco recommends distributing all Cisco CallManager cluster servers, centralized voice gateways, and centralized hardware resources between multiple physical switches and, if possible, multiple physical locations within the campus.

This distribution of resources ensures that, given a hardware failure (such as a switch or switch line card failure), at least some servers in the cluster are still available to provide telephony services. In addition, some gateways and hardware resources are still available to provide access to the PSTN and to provide auxiliary services.

Besides being physically distributed, these servers, gateways, and hardware resources should be distributed among separate VLANs or subnets so that if a broadcast storm or DoS attack occurs on a particular VLAN not all voice connectivity and services are disrupted.

For more detailed information about network infrastructure requirements for a highly available, fault-tolerant campus network, see Chapter 3, "MPLS-Based VPN MAN Reference Topology" and the section QoS for Critical Applications in Chapter 4, "Implementing Advanced Features on MPLS-Based VPNs." Also see the appropriate Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Network Services

After a highly available, fault-tolerant, multi-layer campus network has been built, network services required for IP Communications such as DNS, DHCP, TFTP, and NTP can be deployed.

It is beyond the scope of this document to address these services. For detailed design and deployment guidance for these services, see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Media Resources

A media resource is a software- or hardware-based entity that performs media processing functions on the data streams to which it is connected. Media processing functions include:

- Mixing multiple streams to create one output stream (conferencing)
- Passing the stream from one connection to another (media termination point)
- Converting the data stream from one compression type to another (transcoding)
- Echo cancellation
- Signaling
- Termination of a voice stream from a TDM circuit (coding/decoding)
- Packetization of a stream
- Streaming audio (annunciation)

Note Music on hold is discussed in the following section.

There are basically no design differences with deploying media resources in the self-managed MPLS network. Media resources are normally implemented on the server farm access layer as described in Chapter 2, "Technology Overview."

For more details on media resource design, see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Music on Hold

Music on hold (MoH) is an integral feature of the Cisco IP Telephony system that provides music to callers when their call is placed on hold, transferred, parked, or added to an ad-hoc conference.

Implementing MoH is relatively simple, but requires a basic understanding of unicast and multicast traffic, MoH call flows, configuration options, server behavior and requirements.

This section describes MoH at a high level. For additional configuration and design details for implementing MoH, see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Deployment Basics of MoH

For callers to hear music while on hold, Cisco CallManager must be configured to support the MoH feature, which requires an MoH server to provide the MoH audio stream sources as well as the Cisco CallManager configured to use the MoH streams provided by the MoH server when a call is placed on hold.

The integrated MoH feature allows users to place on-net and off-net users on hold with music streamed from a streaming source. This source makes music available to any on-net or off-net device placed on hold. On-net devices include station devices and applications placed on hold, consult hold, or park hold by an IVR or call distributor. Off-net users include those connected through Media Gateway Control Protocol (MGCP) and H.323 gateways. The MoH feature is also available for plain old telephone service (POTS) phones connected to the Cisco IP network through Foreign Exchange Station (FXS) ports.

The integrated MoH feature includes media server, database administration, call control, media resource manager, and media control functional areas. The MoH server provides the music resources and streams.

You can configure the MoH feature via the Cisco CallManager Administration interface. When an end device or feature places a call on hold, Cisco CallManager connects the held device to an MoH media resource. Essentially, Cisco CallManager instructs the end device to establish a connection to the MoH server. When the held device is retrieved, it disconnects from the MoH resource and resumes normal activity.

Unicast and Multicast MoH

Cisco CallManager supports unicast and multicast MoH transport mechanisms.

Unicast MoH consists of streams sent directly from the MoH server to the endpoint requesting an MoH audio stream. A unicast MoH stream is a point-to-point, one-way audio Real-Time Transport Protocol (RTP) stream between the server and the endpoint device.

Unicast MoH uses a separate source stream for each user or connection. As more endpoint devices go on hold via a user or network event, the number of MoH streams increases. Hence if twenty devices are on hold, then twenty streams of RTP traffic are generated over the network between the server and these endpoint devices. These additional MoH streams can potentially have a negative effect on server CPU resources, network throughput, and bandwidth. However unicast MoH can be extremely useful in those networks where multicast is not enabled or where devices are not capable of multicast, thereby still allowing an administrator to take advantage of the MoH feature.

Multicast MoH consists of streams sent from the MoH server to a multicast group IP address that endpoints requesting an MoH audio stream can join as needed. The self-managed MPLS MAN architecture supports multicast as described in Multicast, page 6-31 and is the preferred design method in deploying MoH in this architecture.

A multicast MoH stream is a point-to-multipoint, one-way audio RTP stream between the MoH server and the multicast group IP address. Multicast MoH conserves system resources and bandwidth because it enables multiple users to use the same audio source stream to provide MoH. Hence if twenty devices are on hold, then potentially only a single stream of RTP traffic is generated over the network.

For this reason, multicast is an extremely attractive technology for the deployment of a service such as MoH because it greatly reduces the CPU impact on the source device and also greatly reduces the bandwidth consumption for delivery over common paths. However multicast MoH can be problematic in situations where a network is not enabled for multicast or where the endpoint devices are not capable of handling multicast.

For information about IP multicast network design, see Multicast, page 6-31.

Recommended Unicast/Multicast Gateways

The following recommended gateways support both unicast and multicast MoH:

- Cisco 6624 and 6608 gateway modules with MGCP and Cisco CallManager Release 3.3(3) or later
- Cisco Communication Media Module (CMM) with MGCP or H.323 and Cisco CallManager Release 4.0, Cisco IOS Release 12.2(13)ZP3 or later, and Catalyst OS Release 8.1(1) or later

- Cisco 2600, 2800, 3600, 3700, and 3800 Series Routers with MGCP or H.323 and Cisco IOS Release 12.2(8)T or later

MoH and QoS

Convergence of data and voice on a single network requires adequate QoS to ensure that time-sensitive and critical real-time applications such as voice are not delayed or dropped. To ensure proper QoS for voice traffic, the streams must be marked, classified, and queued as they enter and traverse the network to give the voice streams preferential treatment over less critical traffic. MoH servers automatically mark audio stream traffic the same as voice bearer traffic with a Differentiated Services Code Point (DSCP) of EF (ToS of 0xB8). Therefore as long as QoS is properly configured on the network, MoH streams receive the same classification and priority queueing treatment as voice RTP media traffic.

Call Processing

Call processing is a critical component of IP Communications design. There are no changes in call processing design when deploying CallManager in a self-managed MPLS environment versus traditional enterprise design other than the deployment models that are described in IP Telephony Deployment Models over the Self-Managed MPLS MAN, page 6-8.

For detailed design guidance for IP Communications call processing, see the Enterprise IP Telephony SRND at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

Cisco Unity Messaging Design

This section focuses at a high level on the design aspects of integrating Cisco Unity with Cisco CallManager in the various deployment models over the self-managed MPLS MAN. The design topics covered in this section apply to both voicemail and unified messaging deployments.

This section does not discuss the full design details of integrating Unity with CallManager because this is fully documented in the IP Communications call processing section of the Enterprise IP Telephony SRND available at the following URL: http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_implementation_design_guide_book09186a008044714e.html.

For additional design information about Cisco Unity, including integrations with other non-Cisco messaging systems, see the Cisco Unity Design Guide at the following URL: http://www.cisco.com/en/US/partner/products/sw/voicesw/ps2237/products_implementation_design_guide_book09186a008022f63b.html.

Messaging Deployment Models

Cisco Unity supports three primary messaging deployment models in the self-managed MPLS MAN:

- Single-site messaging
- Multi-site deployment with centralized messaging
- Multi-site deployment with distributed messaging

Deployments involving both Cisco CallManager and Cisco Unity use one call processing model for Cisco CallManager and one messaging model for Cisco Unity. The messaging deployment model is independent of the type of call processing model deployed.

In addition to the three messaging deployment models, Cisco Unity also supports messaging failover.

All messaging deployment models support both voicemail and unified messaging installations.

Single-Site Messaging

In this model, the messaging systems and messaging infrastructure components are all located at the same site, on the same highly available LAN. The site can be either a single site or a campus site interconnected via high-speed metropolitan area networks (self-managed MPLS MAN). All clients of the messaging system are also located at the single (or campus) site. The key distinguishing feature of this model is that there are no remote clients across the WAN.

Centralized Messaging

In this model, similar to the single-site model, all the messaging system and messaging infrastructure components are located at the same site. The site can be one physical site or a campus site interconnected via high-speed MANs. However unlike the single-site model, centralized messaging clients can be located both locally and remotely across the WAN.

Distributed Messaging

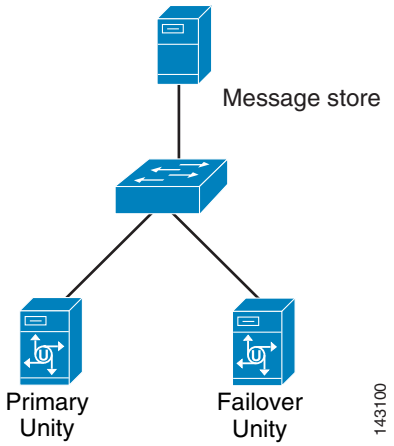
In distributed messaging, the messaging systems and messaging infrastructure components are co-located in a distributed fashion. There can be multiple locations, each with its own messaging system and messaging infrastructure components. All client access is local to each messaging system and the messaging systems share a messaging backbone that spans all locations. Message delivery from the distributed messaging systems occurs via the messaging backbone through a hub-and-spoke type of message routing infrastructure.

No messaging infrastructure components should be separated by a WAN from the messaging system they service. Distributed messaging is essentially a multiple, single-site messaging model with a common messaging backbone.

Messaging Failover

All three messaging deployment models support messaging failover. You can implement local messaging failover as illustrated in Figure 6-12. With local failover, both the primary and secondary Cisco Unity servers are located at the same site on the same highly available LAN.

Figure 6-12 Local Failover of Cisco Unity Messaging



Failover of Cisco Unity Messaging Across the MAN

At present, any configuration that requires failover of Unity across the MAN, such as the Primary Unity server at Data Center 1 and the Failover Unity Server at Data Center 2, requires review by the UBCU TME team.

The UCBU TME team is currently in the process of testing and incorporating Unity Failover across the MAN into the new Unity design guide which will be available shortly.

Cisco Unity and Cisco CallManager support the following combinations of messaging and call processing deployment models:

- Single-site messaging and single-site call processing
- Centralized messaging and centralized call processing
- Distributed messaging and centralized call processing
- Centralized messaging and distributed call processing
- Distributed messaging and distributed call processing

For further details on site classification and a detailed analysis of supported combinations of messaging and call processing deployment models, see the Cisco Unity Design Guide at the following URL: http://www.cisco.com/en/US/partner/products/sw/voicesw/ps2237/products_implementation_design_guide_book09186a008022f63b.html.

Multicast

This section describes how to add multicast VPN as an overlay service on top of an MPLS Layer 3 VPN service. The following major sections are discussed:

- Multicast VPN Service Overview
- Multicast VPN Service Architecture
- Multicast VPN Service Design and Deployment Guidelines
- QoS for mVPN Service
- Multicast VPN Security

- Design Choices for Implementing mVPN
- Multicast VPN Service Management
- Implementing and Configuring the mVPN Service

Multicast VPN Service Overview

The multicast VPN feature in the Cisco IOS software provides the ability to support the multicast feature over an MP-BGP Layer 3 VPN. This service allows users to leverage their infrastructure to deliver multicast with minimal investment.

The Cisco implementation of mVPN offers these benefits:

- The implementation of a data Multicast Distribution Tree (MDT) that allows for a scalable delivery of traffic
- Several PIM options in the core for both data and default MDTs. SSM is offered as an alternative to Anycast RP or Auto-RP
- Support for a broad range of multicast options within the VPN including Anycast RP, Auto-RP, Static RP, Bi-Dir, and accept-register filters in the VPN
- Support for Rendezvous Point (RP) on a PE on a per-VRF basis, VRF awareness of multicast related MIBs, the Cisco mVPN MIB, and MSDP in the VPN (for RP redundancy and RP management service)

Multicast VPN Service Architecture

Service Components

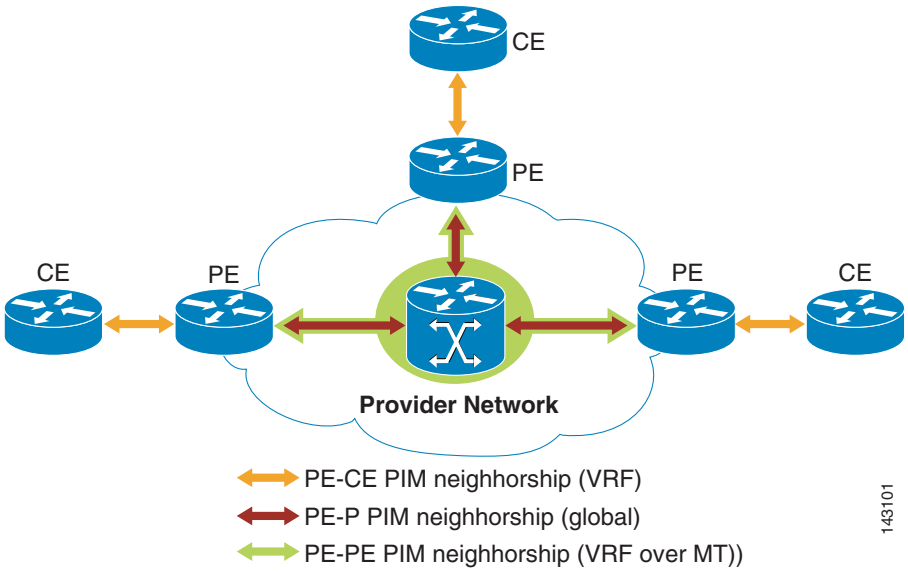
Multicast VPN (mVPN) is an overlay service to the MPLS Layer 3 VPN service that provides multicast forwarding support to VPN users. However mVPN does *not* depend on MPLS. An IP multicast-enabled core is the only requirement. The mVPN feature needs multiprotocol BGP (MP-BGP) VPN support. Therefore mVPN can be implemented as an overlay service on any VPN built with MP-BGP or even without any unicast VPN service.

The following are the required functionalities for the mVPN service:

- Underlying IP routing in the core and in the VPN
- Multicast support in the core
- Underlying MP-BGP-based VPN service
- Multicast VRF support on the PEs and MTI
- Multicast support in the VPN (PE interfaces facing the CE)

Figure 6-13 shows the mVPN service architecture.

Figure 6-13 mVPN Service Architecture



The mVPN service allows for the support of the user multicast traffic in an MP-BGP VPN environment; therefore allowing for support of multicast video, voice, and data within the VPN. The current Cisco implementation of multicast VPN supports:

- Multiple user-facing multicast protocol options (PIM-SM, PIM-DM, SSM, PIM-Bidir, BSR, static or auto RP, and SPT threshold values)
- Multiple multicast core options for the MDT support (PIM-SM, SSM, PIM-Bi-Dir)
- Optimization of bandwidth in the core using configurable data MDTs

The following subsections describe the different service components that should be implemented on the network architecture to provide the mVPN service.

Multiprotocol BGP

Multiprotocol Border Gateway Protocol (MP-BGP) offers a method for network administrators to distinguish which prefixes to use for performing multicast reverse path forwarding (RPF) checks. The RPF check is fundamental in establishing multicast forwarding trees and moving multicast content successfully from source to receiver(s). MP-BGP is based on RFC 2283 multiprotocol extensions for BGP-4.

MP-BGP is also used to propagate the MDT information that allows the PIM neighborships between PE routers to be created and to set up the multicast tunnels for the VPN. The MDT group address is carried as a VPNv4 address along with MDT source information (BGP peering source address).

New Extended Community Attribute

A new extended community attribute has been created to carry MDT source information along with the MDT group address in the BGP updates.

MVRF

Multicast VPN routing and forwarding instance (MVRF) is a virtual forwarding table that supports multicast entries. This feature needs to be enabled on the router for the VRF to support the mVPN service.

Multicast Tunnel Interface (MTI)

Each MDT is associated with an MVRF and the interface that points to this MDT (from the MVRF) is an MTI. The MTI is listed in the outgoing interface list (OIL) for the VRF multicast-group entries. MTI is a virtual interface that gets automatically created when the default MDT is configured in the VRF context.

Multicast Domain (MD)

An MD is a collection of MVRFs that can exchange multicast traffic. Effectively, every PE router in a multicast domain becomes both a sender and a member of the corresponding MDT group.

Multicast Distribution Tree (MDT)

MDT groups are used to encapsulate and transport VPN traffic within the corresponding multicast domain. Multicast routing states for MDT groups are created and maintained in the global routing table only. To a P router, an MDT group appears the same as any other multicast address. It does not require any special handling for routing and forwarding. However because VPN packets need to be encapsulated and decapsulated as they enter and exit (respectively) the multicast domains, a PE router handles MDT groups differently.

A mesh of point-to-multipoint multicast tunnels are established for each multicast domain. At the root of the tunnel is a PE router. The leaves are also PE routers that have MVRFs in the same multicast domain. Typically, there is a single MVRF in a MDT.

There are two types of MDTs:

- Default MDT—Tree created by the mVPN configuration. The default MDT is used for user control plane and low rate data plane traffic, as well as any dense-mode traffic. It connects all of the PE routers with the MVRF in a particular MD. A default MDT exists in every mVPN whether or not there is an active source in the user network.
- Data MDT—Tree created dynamically by the existence of active sources in the user network sending to active receivers located behind separate PE routers. The tree is triggered by a traffic-rate threshold. Data MDT connects only PE routers that have active sources or receivers of traffic. The data MDT is created only by (S,G) Sparse or SSM state in the user mVPN network and is never created for dense-mode (S,G) state.

Multicast VPN Service Design and Deployment Guidelines

The following features are required to support the mVPN service:

- Core and PEs (on the core-facing and loopback interfaces) need to be multicast-enabled (PIM-SM/SSM/BIDIR) and support the different RP mapping options when applicable
- PEs need MP-BGP support for multicast
- PE needs to support GRE and multicast tunnel interface (MTI) for MVRF binding

- PE edge-facing interfaces need to support the user multicast functionalities (PIM type and RP mapping options or the SSM or BiDir range defined)

The following additional features are highly recommended for performance, scalability, and user-specific requirements:

- Support for RR functionality to distribute MP-BGP multicast-related information between PEs
- Support for core optimization using data MDTs
- Hardware-accelerated forwarding
- Management support—mVPN MIB and VRF-awareness of relevant multicast MIBs

Note SSM support in the core is recommended. The following address range is the Internet SSM space: 232.0.0.0 through 232.255.255.255. If SSM is deployed in the core, it is recommended to extend this space into some private SSM space (part of administratively scoped addresses: 239.0.0.0-239.255.255.255); thus the Cisco recommendation is 239.232.X.X. This requires that the **ip pim ssm range** command *not* use the default keyword, because this uses the 232.0.0.0 address space for the data MDTs.

Service Deployment

Enabling mVPN on PE Devices that Support VPN Services

To overlay a multicast VPN service on top of an existing VPN service, the following steps are required:

Step 1 Choosing the PIM mode for the core network.

Cisco recommends PIM-SSM as the protocol in the core. For source discovery, a new attribute is added to BGP , so no additional configuration is required.

A new RD type is used to advertise the source of the MDT together with the MDT group address. PIM SM has been the most widely deployed multicast protocol and has been used for both sparsely and densely populated application requirements. PIM SSM, although newer, is based on PIM SM without an RP. In SSM receivers immediately join the Shortest Path Tree back to the source without joining the RP shared tree. Either PIM SSM or PIM SM are suitable for the default MDT or the data MDT.

PIM SSM is simpler to deploy than PIM SM. It does not require a Rendezvous Point and the core network is a known and stable group of PE multicast devices. Cisco recommends the use of PIM SSM for mVPN core deployments.

Step 2 Choosing the VPN group addresses used inside the core network.

In Step 1, PIM-SSM was selected. The default PIM-SSM range is 232/8; however this address range is designed for global use in the Internet. For use within a private domain, the use of an address out of the administratively scoped multicast range, 239/8, is recommended (RFC2365). Using this private address range makes it simpler to filter on boundary routers.

Cisco recommends using 239.232/16, as this address range is easily recognizable as both a private address and a SSM address by using 232 in the second octet.

In the design discussed in this document, the range is divided for default-MDT and data-MDT. Data-MDT is discussed elsewhere in this document.

Default-MDTs uses 239.232.0.0-239.232.0.256 and Data-MDTs use 239.232.1.0-239.232.255.255. This address range provide support for up to 255 MVRFs per PE router.

Step 3 Configuring the core network for PIM-SSM.

The following commands need to be configured to enable a basic PIM-SSM service.

a. On all P and PE routers configure globally:

```
ip multicast-routing
access-list 1 permit 239.232.0.0 0.0.255.255
ip pim ssm range 1
```

b. On all P interfaces and PE interfaces facing the core configure:

```
ip pim sparse-mode
```

c. On the PE routers configure on the loopback interface used to source the BGP session

```
ip pim sparse-mode
```

Step 4 Configuring the MDT on the VRF.

a. To configure multicast routing on the VRF, configure on all PE routers for the VRF:

```
ip vrf <vrf-name>
mdt default 239.232.0.1 (I just don't like 0.0)
mdt data 239.232.X.0 0.0.0.15 threshold 1
```

b. Choose a unique data MDT address (X) for each VRF that supports IP multicast.

To enable multicast routing for this VRF, configure the following:

```
ip multicast-routing vrf <vrf-name>
```

c. Choosing the PIM Mode for the VPN:

The PIM mode inside the VPN depends on what the VPN edge network is using. Cisco provides automatic discovery of Sparse Mode RPs inside a VPN via auto-rp or bsr, which requires no additional configuration. Static RPs, Anycast RPs, and Dense Mode PIM are also supported on a per MVRF basis.

Optionally, a PE router can be configured as the RP on a per VRF basis.

On the PE, define the RP used within the VPN:

```
ip pim vrf <vrf-name> rp-address <ip address> override
```

Step 5 Configuring the PIM mode inside the VPN.

Configure all PE-CE interfaces for sparse-dense-mode that ensures that either auto-rp or bsr messages are received and forwarded allowing the PE to learn the group to RP mapping inside the VPN. Use sparse-mode for static RP or Anycast RP deployments.

a. Configure on all edge facing interfaces:

```
ip pim sparse-dense-mode
```

Or

```
ip pim sparse-mode
```

The following is a sample of an edge-facing network configuration:

```
interface Ethernet0/0
 ip vrf forwarding <vrf-name>
 ip address 20.0.1.1 255.255.255.0
 ip pim sparse-mode
!
```

The following is a configuration example:

```
ip vrf blue-data
 rd 10:105
 route-target export 10:105
 route-target import 10:105
 mdt default 239.232. 0.10
 mdt data 239.232.10.0 0.0.0.15 threshold 1
ip multicast-routing distributed
ip multicast-routing vrf blue-data distributed

ip pim vrf blue-data rp-address 1.1.1.11 override
ip pim vrf blue-data ssm range 1
access-list 1 permit 239.232.0.0 0.0.255.255
```

The format of the default MDT utilizes 239 as the first octet as a private address, 232 as the second octet to identify it as using PIM SSM, while the third octet can be used to relate the MDT to the VRF that it supports to provide for ease of troubleshooting.

Multicast Core Configuration—Default and Data MDT Options

Selecting Protocol for the MDTs

The selection of the multicast protocol (PIM mode) for the MDTs can be done using access lists that can be configured when defining default and data MDTs in the MVRFs. ACLs define the address range to which a given multicast protocol is assigned. Therefore, depending on the VPN user requirements, different protocol options can be used for each MDT. To minimize complexity, it is recommended that the same PIM deployment options be used for all mVPN MDTs.

Protocol Design Options for Default MDT

Generally, Cisco recommends that SSM be implemented in the MPLS core for all MDTs. SSM allows for a simpler deployment, avoids any RP administration or management, eliminates the RP as potential point of failure, and offers efficiency (does not depend on the RP or shared trees).

In the cases where the users might want to keep their existing multicast configurations such as PIM-SM, using Anycast RP, Auto-Rp, or static RP, they can implement any PIM mode for the default and data MDTs. The support for Bidir in the core also allows for simplicity of the control plane (no SPTs) and scalability (eliminates all [S,G] mroute states in the core because it only uses [*,G]), which makes PIM-BiDir an ideal choice for the default MDT as the default MDT carries all control traffic between PEs. BiDir makes sense for very large scale mVPN deployments to limit the multicast route state maintained in the MPLS core of the network.

Protocol Design Options for RP Positioning

In the case of PIM-SM and BiDir, at least one RP is required in the core multicast network for the MDT. Because traffic traverses the RP shared tree (*,G), the RP should be placed in a centralized location and should not be placed on a PE to avoid additional load on the edge routers.

Addressing Ranges and Considerations

Multicast Core—MDT Addressing

Each multicast domain is assigned a distinct group address from a pool of multicast addresses. The group ranges reserved for multicast domains are called MDT groups in this document. The recommended group range for multicast domains is a subset of the administratively scoped addresses: 239/8. The administratively scoped range of addresses is a private address space that is not meant to be exchanged on the Internet.

Note that MDT group addresses are global in the core; each VPN has a unique set of MDT addresses (default and data). Therefore, the addresses cannot overlap if PIM-SM is used.

There is a specific range for Internet SSM (232.0.0.0 to 232.255.255.255). It is not recommended to select from these addresses for SSM default and data MDTs because they are public Internet SSM group addresses. Cisco suggests using 239.232/16 for private SSM addresses. Note that the choice of SSM for MDTs in the core allows optimization of the address space as a given group address may be used by several sources. With SSM each PE MDT has a unique (S,G) even if the same group address is used by different PEs.

When configuring the mVPN, ensure that all PE routers of a given VPN have the same default MDT group.

The setup of MDT and creation of the MTI are conditioned by certain addressing parameters for the MP-BGP update source and PIM neighbor addresses. It is important to set the proper addressing for the mVPN to be up:

- The MTI takes its properties from the interface used for the BGP peering (MP-BGP), usually defined as a loopback address using the *update-source* parameter of BGP.
- The RPF check done for the PIM neighborship also uses the BGP information (next hop) as no unicast routing protocol runs on the MTI. Therefore the PIM peer ID and BPG next-hop addresses of a PE must match. Note that this next hop must be reachable by the core P routers for RPF check.
- The BGP peering address is also used as the source address of the local PE for the multicast tunnels (root of the default and data MDTs).

VPN Network Multicast Addressing

There are no restrictions on the addressing within a VPN because a private address space is provided per definition of the service. VPN address spaces may overlap unless extranets are configured or Internet access is provided.

Caveats

Some other restrictions in addition to the platform-specific issues mentioned in the sections above include:

- If the core multicast routing is using source-specific multicast (SSM), then the data and the default MDT groups must be configured within the SSM range of addresses. Cisco recommends using a range of 239.232/16.
- Data MDTs are not created for VRF PIM dense-mode multicast streams. PIM DM relies on a tree that reaches all CE routers; therefore, the default MDT is used.

- Multiple BGP update sources are not supported and configuring them can break the mVPN RPF check. The source IP address of the mVPN tunnels is determined by the highest IP address used for the BGP peering update source. If this IP address is not the IP address used as the BGP peering address with the remote PE router, mVPN does not function correctly.

QoS for mVPN Service

Because the mVPN feature is generally an additional service provided along with Layer 3 VPN unicast, users expect the same QoS and flexibility. This specifically applies to the transport of their time- or business-sensitive multicast traffic that requires prioritized transport and low drop rate. MoH and video streaming are examples of multicast applications that benefit from QoS.

The QoS requirements include:

- The existing QoS transparency models (pipe, short pipe, or uniform) for unicast MPLS VPN traffic should be enhanced to support mVPN traffic.
- All the ingress and egress QoS policies on all the CE, PE, and P routers should be modified to support multicast or mVPN traffic.
- Statistical reporting and measurements need to provide accurate information for mVPN traffic.

Implementing QoS for the mVPN

Baseline IP performance degradation or failure naturally impact the quality of service offered for mVPN. Note that the mVPN traffic is GRE or IP in IP traffic sent over the multicast core. Therefore it is a good practice to ensure initial good performance and reliability of the IP core and the mVPN service itself. Other factors that influence QoS include convergence, scalability, and performance.

Implementing QoS for mVPN on the PE Routers

PE routers are the devices that are the most critical points in terms of performance and are also the entry point for the multicast VPN user traffic. Therefore most aspects of QoS implementation are recommended on the PEs, depending on platform support.

Recall that the PEs perform most of the service features (mVPN encapsulation, decapsulation, and lookups), keep the VPN multicast information (MVRF, mstates in the global routing table for data and default MDTs), and cumulate this information with regular unicast route and control plane information. The PE routers are also the point where other services are implemented (for example, security and 6PE functionalities).

Implementing QoS for mVPN on the P Routers

QoS support in the core implies enabling QoS in the core for IPMC traffic, as the multicast VPN packets are sent as IP-in-IP or GRE-encapsulated multicast-traffic over the MPLS core.

Implementing QoS for mVPN on the CE Routers

The CE could be the place in the network where classification and potential marking or remarking of packets would be done. This offloads the PE. If this is not done, then the multicast packets arrive at the ingress PE interfaces marked according to the user classes of services and might need policing, class mapping, and remarking.

Multicast VPN Security

The VPN multicast feature in itself is a secure service ensuring the transport of multicast information. Per definition, a VPN service must guarantee privacy of the user traffic. Security mechanisms should be implemented from the very beginning of a service offering because they are critical to service definition, assurance, and SLAs fulfillment.

Security of the mVPN service can be considered at two levels:

- Protection of resources—This ensures that a mVPN service is up and functioning; i.e., transporting the user information. The resources to be protected are the CPU load and memory usage of routers (Ps, PEs, CEs, RRs), link bandwidth protection and guarantee, and resource access protection (physical and administrative-login). The resources of the core and the edge are shared by several mVPN users; therefore a node or link failure would have an impact on the service of several users.
- Protection of the user information (privacy, integrity)—This means protecting the access to the control and information flows (permission to be a receiver for a group or an MDT, permission to connect to a VPN and receive VPN multicast data,and isolation of flows between different user groups),and control of information sources and access to the VPN (on the PE-CE and access to the MDT resource).

The different security design recommendations for the multicast VPN service are further divided according to the following four types of features, each of which might present resource protection and user information security levels:

- Access control—Which sources may send multicast data (in the VPN or in the core)
- Admission control—Who receives the multicast data (receivers in the VPN or MDT members)
- Data plane control—Protecting the data exchanged over the mVPN service
- Control plane control—Protecting the control information

Implementing Security for mVPN Service

Several important levels of security should be implemented on the PE, which is an important element of mVPN service security because it is located between VPNs and the MPLS core network and performs the most mVPN-relevant operations. Therefore it is exposed to a high control plane and data traffic load. Regarding resource protection, you want to limit the control plane memory and CPU usage (number of VPN multicast routes, cycles to generate trees, cycles to maintain BGP information, and so on) as well as the data plane load (unicast and multicast traffic forwarding, double lookup, and encapsulation or decapsulation of the mVPN traffic). It is important to control access to the multicast information flows and the control plane itself to avoid intrusions, protect integrity, and to respect the separation of the traffic between mVPN service users.

In the case of resource protection, the mVPN service requires a careful design of the core in terms of the number of MDTs and multicast VPN routes. These factors impact the load on the P routers (amount of mstates and traffic load). When using PIM-SM or BiDir in the core, the RP functionality and placement should be carefully considered for redundancy, access control (that is, which PE accesses an MDT as a source/receiver), and control plane load.

Levels of security that can be implemented on the CE router include:

- Router access
- Control and limitation of the multicast data sent to the PE, which reduces load and risks of DoS attacks and ensures core and PE scalability
- Limitation of the resources used by the control plane on the CE

Care should be taken when configuring the RP parameters to ensure controlled source access and receiver access to the VPN multicast groups, to restrict control plane load because of the user-facing RP functionality, and to implement redundancy for the RP service (RP service feature assurance) because the PE is the most exposed component of the mVPN service.

Note that all unicast Layer 3 VPN security recommendations are also relevant in the case of an mVPN service; the PE and core are shared by unicast and multicast VPN traffic because both types of traffic are transported on the same infrastructure.

Access Control

Source access is an important security aspect of multicast in general and applies to an mVPN service as well for two reasons: it affects the privacy guarantee of a VPN and it is a resource protection measure, protecting the user resources as well as the core and edge multicast resources. A rogue source can generate a lot of traffic and overwhelm receiver and network resources along the trees, which can be a result of DoS attacks.

PIM-SM and BiDir

PIM-SM and BiDir use RPs for source registration. The following features allow access control of sources registering either at the multicast core RP or a VPN user network RP level:

- Source registration

  To define access lists and route maps to filter incoming register messages based on the source address or (S,G) pair, use the source registration feature configured on the RP themselves. This feature defines which sources or (source, group) parameters are accepted while the rest are rejected. It is a global or VRF aware command (**ip pim accept-register [list <acl>][route-map <>map]**). Use this feature when the sources and multicast groups are known and when the RP is close to the sources.

  Certain considerations should be taken into account. The first hop routers located between the source and the RP still receive the traffic from any source. As a result, this traffic creates control and data workload on these routers as they process packets and generate corresponding mstates. The source-register filter within a VPN can be used to limit which traffic is accepted as multicast VPN traffic; therefore limit the traffic across the multicast core (on the MDT). In the core, because the network knows the MDT group addresses to expect, this feature allows unexpected multicast group traffic (MDT or none) over the multicast core to b e denied. For example, imagine that a rogue device sending traffic from a source to an MDT group, thereby becoming an illegitimate source on the VPN. This is possible with multicast protocols such as PIM-SM and BiDir where an RP is being used. A source sending to the RP is requested to only have a PIM neighborship with the RP and to register with the RP. This is possible because PIM does not currently have any authentication mechanisms. One way to prevent this from happening is to set appropriate source filters on the RP as described above.

- Local loopback RP

  An alternative to source access control when using RPs is to implement the local loopback RP method that is used to blackhole locally the traffic sent by rogue sources towards a multicast group (**ip pim rp-address <local_loopback> <acl>**). This command requires the user to explicitly define which group addresses are not authorized. This is less restrictive than the previous option in the sense that an undefined source is able to send if not listed in the access list. The feature has the advantage to avoid rogue traffic even to the first hop routers located on the path between local router and the RP. This avoids control workload on these routers but requires local configuration on each router and cannot prevent load on the very first router and multicast to local sources. This is more efficient in terms of first hop routers protection; however the above limitations should be kept in mind.

- **ip pim accept-rp**

  The **ip pim accept-rp** is used by routers to select which RPs are accepted for a given set of multicast groups. This feature can be applied within the VPN and in the core. In the core it prevents a rogue RP from forwarding information on the MDTs in the cases of PIM SM and BiDir in the core. Note that the MDT group addresses are well known because they are predefined (default MDT address plus a range of addresses for data MDTs). Therefore it is simple to define the groups for which a given RP set is accepted or rejected.

The previous features emphasize that previous knowledge of VPN user traffic patterns and a proper addressing scheme in the core are important design recommendations when considering access control security. See Addressing Ranges and Considerations, page 6-38 for addressing recommendations.

PIM-SSM

In the case of SSM, which does not use RPs, source control is implicitly implemented because receivers must have a prior knowledge of a source to request for their local router to join the corresponding channel (use of [S,G] versus [*,G]). This presents a certain level of protection against the potential threat of DoS and VPN privacy attacks from rogue sources and can be used in the VPN user network (PE-CE and beyond) as well as in the core.

Access Lists

Any mVPN packet belonging to a user multicast flow is sent site-to-site over the default or data MDT. The VPN user network sources are thus effectively sources of multicast core traffic. In the absence of filtering everything is flooded over the default MDT or sent over the data MDT. Even if data MDT creation and access can be restricted, traffic is sent over the default MDT and affects the core. The following two practices are recommended to protect the multicast core:

- ACLs on the PE-CE should be used to limit the VPN user multicast traffic that travels site-to-site over the MDTs. This preserves the core resources, avoids some DoS attacks, and preserves PE resources by avoiding unnecessary processing of multicast traffic. Some knowledge of VPN user sources and group addresses is necessary to implement these ACLs. Using data ingress ACLs is highly recommended because it preserves network resources as well as PE resources.

Note If ACLs are configured on the PE-CE link for security in conjunction with the mVPN feature, a decrease in performance load might be noticed because of the scanning of packets.

- Implementation of QoS can provide core resources protection because traffic policing and to some extent traffic shaping and congestion avoidance can limit the traffic sent over MDTs (policing based on VPN user addressing at the ingress, WRED and shaping based on MDT group address at the egress). In these cases, the use of QoS allows one to protect the core resources.

Admission Control

Receiver access to the multicast traffic is an important VPN flow privacy concern. Within the VPN, receiver access to a multicast group or to the traffic sent by a specific source for a group can be limited using IGMP access groups. This is a security feature to be implemented within the VPN. It is configured on the PE if any host is expected to connect directly to the PE router.

Regarding resource protection on the PE, use of the IGMP access groups limits the load of the PE to a certain extent because a PE subscribes only to a data MDT if there is a receiver requesting the corresponding (S,G) traffic.

The join of a rogue P to the MDT tree as a leaf only necessitates a PIM neighborship with the RP or one of the core routers. Because PIM does not currently support peers authentication, this is a potential security issue for the mVPN service. However note that this necessitates physical connectivity access to one of the core or edge routers.

Control Plane Protection

For control plane, physical and login access to the devices (PEs, CEs, Ps, and RRs) and links should be protected. Also keep in mind that multicast uses unicast routing information; therefore security of the control plane at this level should also be considered.

BGP Neighbors

Existing BGP neighbor authentication mechanisms (MD5 authentication) for PE-PE or PE-RR information-exchange can be used.

PIM Neighbors

Overall, an intelligent design of the PIM options (see below for RP choices) for the core can potentially reduce the risks of attacks or failure, therefore providing a good resource protection design, redundancy (RRs, RPs, and sink RP in auto-RP cases), and careful selection of thresholds (SPT). The **ip pim neighbor-filter** *acl* command allows you to administratively deny a PIM neighbor from participating in PIM and is a first level of protection. To set up the PE-to-PE PIM adjacencies, the PIM router ID must be the same as the BGP peering address, which presents some level of protection against rogue PE routers.

Limiting mroutes

An important factor in terms of mVPN support is the number of mroute states in the core and on the PEs, which is a potential door for DoS attacks. Limitations of the number of routes exchanged among BGP peers and of routes installed per VRF are two ways to limit the impact of this type of attacks:

- Limitation of the number of multicast routes in an MVRF—The **ip multicast [vrf vrf-name] route-limit limit [threshold]** command limits the number of multicast routes that can be added to a multicast routing table, thus countering any attempt to fill up multicast routing states. From a service assurance and monitoring perspective, a threshold can be created and an error message is created when the limit is exceeded. The message recurs until the number of mroutes drops under the limit set by the limit argument.

Note The **max route** option of the VRF setup is applied to unicast routes only and cannot be leveraged for multicast flows.

- Limitation of the number of multicast routes accepted from a BGP neighbor—The **max route** option per-BGP neighbor can be used to limit mroutes received from a neighbor. The following example shows how to configure the maximum value for the IPV4 address family, which applies to IPV4 unicast and multicast routes:

```
pop2-7206-11-PE(config-router-af)#neighbor <IP@> maximum-prefix <MaxPrefixValue>?
<1-100> Threshold value (%) at which to generate a warning msg
restart Restart bgp connection after limit is exceeded
warning-only Only give warning message when limit is exceeded
<cr>
```

Limiting mstates

Limiting the impact of VPN user traffic on the creation of data MDTs triggered by traffic sourced in the VPN network is an important aspect of core resources protection (data MDT creation process is dynamic and requires core CPU and memory resources). The following features achieve this:

- The data MDT command (for example, mdt data 239.232.1.0 0.0.0.255 threshold 50 list 5 of the mdt data MDT_data_add wildcard threshold value_thresh list acl_# command) provides some level of protection because an access list defines the type of VPN user multicast traffic that triggers the creation of a data MDT (an ACL can define a group address or a source and group address). Cisco recommends carefully defining the VPN sources and groups that may trigger the creation of a data MDT. Note that this implies that the core must be configured for support of the multicast sources IP addresses of the user.
- For a given VPN, the amount of MDTs is limited by default to 256 (255 data MDTs and one default MDT) and can be configured to a lower value. Therefore, the impact of having potential sources creating a DoS threat because of high control plane activity and memory usage is limited by this implementation option. The 255 limitation provides some level of protection of core resources from a control plane perspective and should be used together with other resource protection features. However note that DoS attacks can take place because high rate sources send over MDTs and potentially use up multicast core resources.

The choice of data MDT threshold may also be an effective element of PE and P resource protection because its value defines which amount of traffic triggers the creation of a data MDT and therefore restricts the use of control plane resources for the dynamic setup of the MDT and the mroute state maintenance in the core.

As a general good practice with respect to data MDT monitoring, use the following command in vrf mode, which allows data MDT reuse to be logged:

```
Router(config-vrf)# mdt log-reuse
```

Note The feature enabled with the **mroute state limit** command per-interface and per-VRF is currently designed to provide resources protection and prevent potential DoS attacks (for example, the limitations of mroute states received per interface). It is used to limit the number of mroutes accepted from the PIM neighbor.

RP Protection

When RPs are configured in the core or in the VPN, Cisco recommends using redundancy mechanisms, depending on your RP mode, such as:

- Anycast RP using MSDP in the case of static RPs
- Mulitple candidates and mapping agents for auto-RP
- Multiple candidates in case of BSR, as well as redundancy of the bootstrap router

The RP functionality may be CPU-intensive on a router so care should be taken when choosing which routers support this functionality.

If auto-RP is used, use the **ip pim rp-announce-filter rp-list group-list** command on the mapping agent to restrict the routers that are accepted as RP candidates for a specific group range. This feature is recommended in the core and at the VPN network level to filter out false C-RPs. On the RPs themselves, when using MSDP, Cisco recommends using the **ip msdp SA-filter** and **ip msdp SA-limit** features to filter incoming SAs from a peer and to protect against DoS attacks by limiting the global number of SA messages on the local MSDP peer.

Overall, when designing the service and applying security-oriented features, some features might incur additional processing at the control of data plane. The impact of these features on the overall performance of the service should be kept in mind.

Design Choices for Implementing mVPN

Multicast Address Ranges for the MDTs in the Core

To provide resource protection for more scalability and to accommodate several levels of service and types of requirements, it is possible to use multicast address ranges to offer different core multicast support options to different users. For example, for a given VPN user group A that is associated with the VRF mcast1 and with the default MDT address 239.232.0.1, the data and default MDTs traffic is carried in the core using PIM SSM; for user group B that is associated with the VRF mcast2 and with the default MDT address 239.232.0.2, the default MDT uses PIM-BiDir and the data MDT uses SSM. Note that options such as RP mode (static, auto, and BSR) and the RP router choice can be also configured per user group at the MDT level using address ranges.

Data MDT Implementation Options

The following are data MDT implementation options:

- An advantage of using data MDTs is the ability to optimize flows across the P network. This offers core optimization to delivery scalability and an optimized use of core resources (memory and CPU) because it reduces the load on the PEs in an mVPN. If the user has multicast data streams with heavy throughputs or one-to-many type of streams, the use of data MDTs is highly recommended. Data MDTs are triggered at the PE closest to the source. PEs with interested receivers send dynamic “joins” to the data MDT after they receive a receiver IGMP report for the associated group. The setup of the data MDT itself creates additional states in the core and generates CPU load on the PEs. It might not be necessary to set up data MDTs for streams whose sources reach almost all VPN sites (remote PEs) or those with a very low throughput.
- Access to the data MDT—Address ranges can be used to regulate which source multicast addresses may trigger the creation of a data MDT. This is a good practice (remember the trade-off for data MDTs: additional states created in core network for a more efficient bandwidth usage) and it also allows a way to attribute certain group multicast prefixes to certain VPN multicast address prefixes for the data MDT when user stream addresses are well known.
- The value of the data MDT threshold is an important parameter for the core stability and scalability. It can be configured per data MDT group address range in a VPN and customized per-user VPN multicast stream (defined using access lists).

Convergence for the mVPN Service

Dependencies

Multicast VPN does not use MPLS switching in the core. Convergence depends only on routing protocol and multicast parameters. An MPLS-LDP failure does not interrupt multicast traffic forwarding because the mVPN solution does not rely on MPLS.

If a core link fails or is shut down and this link belonged to the path of one or several of the multicast trees, the corresponding traffic (control or data) is affected and some data loss occurs. The traffic is rerouted to a backup path. The delay is mainly because of the IGP convergence time because PIM relies on the unicast routing and triggers a new RPF computation when the IGP topology changes.

Convergence of Default and Data MDTs

Multicast VPN convergence depends on core multicast convergence and MP-BGP convergence characteristics. For in-depth information about multicast, see the convergence subsection in the multicast documentation at the following URL: http://www.cisco.com/en/US/partner/products/sw/iosswrel/ps1829/products_feature_guide09186a00800da1cd.html.

In the case of MP-BGP peer failure or update, connectivity is lost for a single-homed multicast VPN site. Convergence of BGP forces the convergence of the MDTs for sites that depend on more than one PE. See the BGP convergence documents for MP-BGP convergence because it is not different from BGP convergence. Note that the introduction of redundancy for the RRs in the design helps with the availability of MP-BGP.

Convergence of Data MDTs (Establishing a Data MDT)

The data MDTs are dynamically set up after the crossing of a bandwidth threshold by the MDT traffic defined for a given set of sources (ACLs). The statistics to determine whether a multicast stream has exceeded the data MDT threshold are examined once every ten seconds. If multicast distributed switching is configured, the time period can be up to twice as long. The establishment of a data MDT occurs by notification to the PEs of the MD using a message sent to ALL-PIM-ROUTERS (224.0.0.13) on UDP port 3232.

The source PE starts using the data MDT three seconds after sending the notification and also stops using the default MDT. It is possible that some PEs lose some packets of the source multicast flow if they are not already joined to the data MDT at the time the source PE switches over.

Implementing and Configuring the mVPN Service

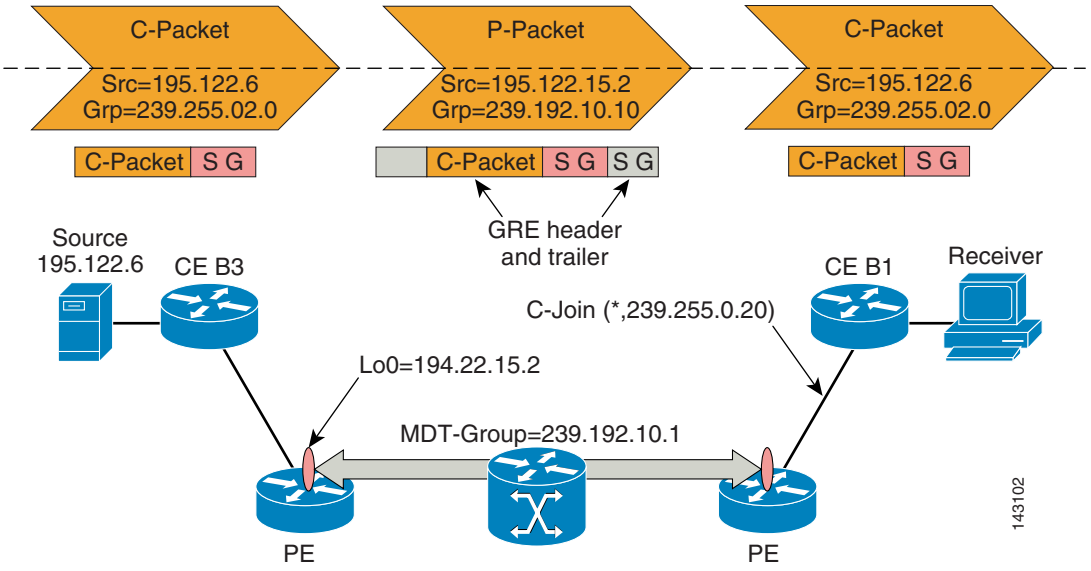
Life of a Packet

Both user control and data traffic (C-packet) is sent over the multicast tunnel. Data traffic for a given multicast group (S,G) or (*, G) is only sent over the multicast tunnel if a remote host registered for the group or if it is a dense-mode user group. The multicast traffic incoming on the PE-CE link is handled as follows:

1. A lookup is done in the MVRF to determine the output interfaces and proceeds to the RPF check. If the tunnel interface is in the outgoing interface list, the packet is encapsulated in GRE (or IP-IP) with the (S, G) information corresponding to S= BGP peering address of the PE router and G = data or default MDT address.
2. The encapsulated multicast packet appears as generated locally by the PE router.
3. A second lookup (including RPF check) is done on this packet in the global multicast routing table. The packet is then forwarded (after appropriate replications) as multicast traffic over the MDT towards the other PE routers. P routers see only P-packets, so they do not build state for traffic inside the VPN. P-packets go to each PE router (except in the case of data MDTs) in the multicast domain. When the P-packet is received from a multicast domain, the PE router duplicates it towards core-facing interfaces if needed. If the MVRF is listed as an entry in the OIF of the global table, it removes the GRE-IP (or IP-IP) encapsulation and then continues to route the packet in the target MVRF identified by the MDT group address in the outer header. The C-packet is sent to the appropriate OIFs according to the MVRF.

Figure 6-14 shows the packet flow including the addressing in the packet headers.

Figure 6-14 Life of a Packet



Configuring the mVPN Service

Use the following commands to enable multicast global and VRF routing:

```
ip multicast-routing [distributed]
ip multicast-routing vrf VRF_name [distributed]
```

Note For cases where MDS is available, use the **distributed** keyword.

Use the following commands to configure the VRF:

```
ip vrf VRF_name
 rd rd
 route-target export rt
 route-target import rt
 mdt default Multicast_Default_MDT_IP@ mdt data <Multicast_Data_MDT_IP@> <range> threshold
threshold list access_list_number
```

Use the following commands to configure the loopback interface for exchanging MP-BGP information:

```
interface Loopback0
 ip address IP_address mask
ip pim sparse-mode
```

All other VRF-linked or core-facing interfaces must be PIM-enabled (sparse-mode or sparse-dense mode, depending on the design options used). Use the following commands to configure PIM on these interfaces: **interface** int_type slot/port

```
ip address address mask ip pim sparse- mode
```

Use the following commands to configure the core multicast, based on the core multicast protocol options:

```
ip pim rp-address IP_address
access_list_number ip pim spt-threshold {infinity | 0 |value}
```

Use the following command to configure SSM in the SSM default range in the core:

```
ip pim ssm range acl-number
           access-list number permit address range
```

Use the following commands to configure the multicast protocol options in the MVRF:

```
ip pim vrf VRF_name rp-address IP_address> access_list_number
ip pim vrf VRF_name bsr-candidate slot/port
ip pim vrf VRF_name rp-candidate slot/port group-list access_list_number
ip pim vrf <VRF_name> ssm range access_list_number
```

For the Cisco 12000, an additional step must be used to enable mVPN:

```
router bgp 1
<snip>
 address-family ipv4 mdt
 neighbor 125.1.125.15 activate
 neighbor 125.1.125.15 send-community extended
 neighbor 125.1.125.16 activate
 neighbor 125.1.125.16 send-community extended
 exit-address-family
```

Use the following commands to enable SNMP and the available MIB notifications:

```
logging source-interface management-interface-slot/port #
syslog server setting logging syslog-server-IP-address
snmp-server community private RW
snmp-server community public RO
snmp-server trap-source management-interface-slot/port
snmp-server location location
snmp-server contact name_or_email
snmp-server enable traps pim neighbor-change rp-mapping-change invalid-pim-message
snmp-server enable traps ipmulticast
snmp-server enable traps msdp
snmp-server enable traps mvpn
snmp-server host SNMP-server-IP-address public
```

For more detailed information, see the following URL:
http://www.cisco.com/en/US/partner/products/sw/iosswrel/ps1839/products_feature_guide09186a0080110be0.html#wp1025098.

Troubleshooting Tips for mVPN

The following are some recommended troubleshooting guidelines:

- Verify the global domain information as follows:
  - Confirm that IP unicast works properly in global domain (IGP and BGP) (multicast uses IP unicast routing information and do not work unless the unicast works correctly).
  - Check PIM neighborships in global domain (PEs-Ps and Ps-Ps).
  - If PIM-SM used, check that an RP is known by all PEs and Ps routers and consistent group—RP mapping exist.
  - Check the mroutes for the default MDT groups on all routers in a path starting with a receiver (make sure to pay attention to flags—check to ensure that the groups shows the correct PIM mode)
- Check the VRF domain information as follows:

- Check unicast VRF tables and neighborships (addresses attached to VRF, route distributions, RTs, RDs, and, if needed, check RR function)
- Check PIM neighborships within VRF (between PEs)
- If PIM-SM used, check that an RP is known by all PE and CE routers consistently (on the VPN user side)
- Check MDT usage (use the **sh ip pim mdt bgp** command)
- Check mroutes on all routers for a path starting with the receiver (make sure group is in correct PIM mode)
- Check traffic flow hop-by-hop

Note Check for the Z flag for default MDT and the Y flag for data MDTs.

Best Practices for Configuring and Troubleshooting mVPNs

The following are some best practices for configuring or troubleshooting mVPNs:

- The update source interface for the Border Gateway Protocol (BGP) peerings must be the same for all BGP peerings configured on the router for the default MDT to be configured properly. If you use a loopback address for BGP peering, then PIM sparse-mode must be enabled on the loopback address.
- The **ip mroute-cache** command must be enabled on the loopback interface used as the BGP peering interface for distributed multicast switching to function on the platforms that support it. Do not enable the **no ip mroute-cache** command on these interfaces.

Useful Show Commands

Table 6-2 lists CLI commands that can be used for troubleshooting mVPN service configuration.

Table 6-2 Useful Show Commands for Troubleshooting mVPN Configuration

Show command	Function
`show ver`	Cisco IOS version—check that mVPN is supported.
`show ip mroute show in`	Interface type of IIF and OIF (Layer 1 and Layer 2; also specify whether sub-interface is used and which type)
`show ip pim mdt`	Number of MVRFs
`show ip pim mdt bgp`	Other PEs multicast-enabled neighbors per MVRF
`show ip pim vrf X neighbor`	Total number of PIM neighbors in each VRF
`show ip pim neighbor"`	Total number of PIM neighbors in global table
`show ip igmp vrf X groups`	Total number of IGMP groups in each VRF
`show ip igmp groups`	Total number of IGMP groups in global domain
`show ip traffic`	Number of PIM/IGMP messages sent/received (itemize by message type)
`show ip mroute vrf x count`	Number of G per MVRF
`show ip mroute vrf x count`	Number of S per G per MVRF
`show ip mroute vrf x`	Number of OIFs per mroute in VRF

Table 6-2 Useful Show Commands for Troubleshooting mVPN Configuration (continued)

`show ip mroute vrf x count`	Memory used for each VRF MRT
`show ip mroute count`	Number of G in global MRT
`show ip mroute count`	Number of S per G in global MRT
`show ip mroute`	Number of OIFs per mroute in global MRT
`show ip mroute count`	Memory used by global MRT
`show mem summary or show proc mem`	Total memory used
`show ip pim vrf x mdt send`	Number of data MDT's sourced
`show ip pim vrf x mdt receive`	Number of data MDT's received
`show proc cpu | exc 0.00`	CPU utilization on RP and LCs
`show ip vrf interfaces` *`vrf_x`*	Interfaces attached to a VRF- Number of CEs per PE per
`show ip vrf` *`vrf_x`*	VRF configuration
`show ip pim mdt bgp`	Visualize the multicast VPN BGP updates
`show run | incl pim show run | incl multicast`	Visualize interface and global PIM and multicast CLIs from the running configuration
`show ip mds interface [vrf` *`vrf-name`*`]`	Display Multicast Distributed Switching (MDS) information for all the interfaces on the line card
`sh ip pim vrf X mdt {send | receive}`	Check the data MDT creation and history
`sh ip mds forwarding`	Check the MDS forwarding information

Ethernet over MPLS

There are requirements within the enterprise network to be able to extend Layer 2 network functionality across a MPLS core to support services such as server clustering, which rely on the ability to use a Layer 2 path between servers, as well as other legacy protocols. For point-to-point Layer 2 connectivity, Ethernet over MPLS (EoMPLS) based on Martini drafts provides this capability.

This section describes how to add EoMPLS as an additional service to MPLS Layer 3 VPNs. The following major sections are included:

- EoMPLS Overview
- EoMPLS Architecture
- Technical Requirements for EoMPLS
- EoMPLS Configuration and Monitoring

EoMPLS Overview

EoMPLS technology leverages an MPLS backbone network to deliver Transparent LAN Services (TLS) based on Ethernet connectivity to the customer site. The concept of Transparent LAN Services is straightforward; it is the ability to connect two geographically-separate Ethernet networks and have the two networks appear as a single logical Ethernet or VLAN domain. Such a VLAN transport capability

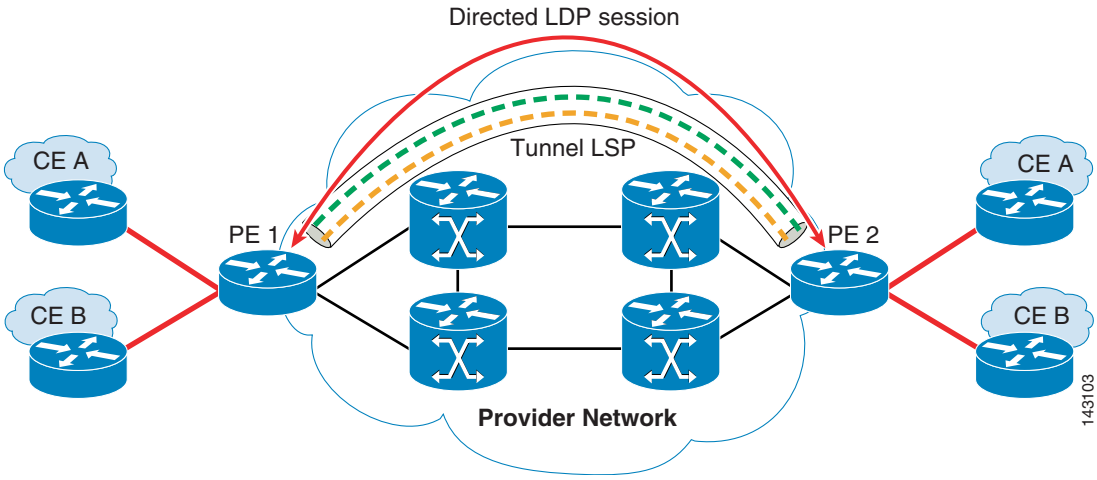
allows customers to deliver a service that allows VLAN networks in different locations within a metro service area to be cost-effectively connected at high transmission speeds, allowing for the ability to carry legacy protocols and provide for Layer 2-based services such as server clustering.

When EoMPLS is deployed in conjunction with MPLS VPN, the network is able to provide tremendous flexibility in the variety of both Layer 2 and Layer 3 network services that can be provisioned over a single, simplified, integrated MPLS backbone network.

EoMPLS Architecture

EoMPLS is a specific implementation of AToM for Ethernet. Figure 6-15 shows the use of a tunnel through an MPLS core network to provide a Layer 2 point-to-point circuit, using AToM, to provide a Layer 2 connection between two separate networks.

Figure 6-15 EoMPLS Architecture



EoMPLS imposition routers must be able to route generic VLAN packets over MPLS backbone. There are four major parts needed to provide EoMPLS service:

- Dynamic MPLS tunnels
- Targeted LDP sessions
- Two-level labeling
- Label imposition/disposition

The EoMPLS implementation is based on the Martini drafts.

A dynamic label tunnel interface is created between the two EoMPLS imposition routers (EoMPLS PEs), an LDP session is formed over this tunnel and MPLS labels are exchanged for each VC to be routed over this tunnel.

Two-level labeling is then used to transport the Layer 2 packets across the backbone. The top label is used to traverse the backbone. The bottom label is the only label seen by the egress LER and is used to identify the VC and the corresponding egress VLAN Layer 2 interface.

EoMPLS has several characteristics that the customer needs to understand to make effective use of this technology:

- Establishing an EoMPLS circuit requires that the edge network be assigned a specific physical port on an LER device, such as a Cisco 7600. The identification of that physical port is a critical element in the binding of the MPLS Label assigned to the customers EoMPLS VC.
- A customer may have more than one EoMPLS VC per physical port as long as the Ethernet traffic transmitted from the customer site to the PE device can be distinguished as having specific 802.1Q headers for each EoMPLS VC by the PE device.
- EoMPLS VCs are point-to-point transmissions only, as explicitly specified in the IETF Martini draft specifications.
- Traffic sent between the imposition/disposition routers (between LERs) over an EoMPLS VC take the same path across the IP/MPLS backbone. The LSP may change because of routing changes inside the MPLS core network.
- Adding/removing a point-to-point Layer 2 VC requires configuration of the two VC endpoints (at the two LERs).
- Provisioning a VC involves defining an endpoint of a Layer 2 VC at each of the VLAN interfaces at the PE router on the interface that connects to the CE.
- The two LERs at the ingress/egress points of the IP/MPLS backbone (the PE routers) are the only routers with knowledge of the Layer 2 transport VCs. All other LSRs have no table entries for the Layer 2 transport VCs. This means that only the PEs require software with EoMPLS functionality.

MPLS VC Circuit Setup

A virtual circuit is an LSP tunnel between the ingress and egress PE, which consists of two LSPs because a uni-directional LSP is required to transport Layer 2 PDUs in each direction. A two-level label stack, where the Level 1 label is the VC label and the Level 2 label is the VLAN tunnel label, is used to switch packets back and forth between the ingress and egress PE.

The VC label is provided to the ingress PE by the egress PE of a particular LSP to direct traffic to a particular egress interface on the egress PE. A VC label is assigned by the egress PE during the VC setup and represents the binding between the egress interface and a given VC ID. A VC is identified by a unique and configurable VC ID that is recognized by both the ingress and egress PE. During a VC setup, the ingress and egress PE exchange VC label bindings for the specified VC ID. The VC setup procedures are transport-independent. The detailed VC setup procedure occurs as follows:

1. An MPLS Layer 2 transport route is entered on the ingress interface on PE1.
2. PE1 starts a remote LDP session with PE2 if none already exists. Both PEs receive LDP KeepAlive messages from each other, reach OPERATIONAL state, and are ready to exchange label bindings.
3. The physical layer of the ingress interface on PE1 comes up. PE1 realizes there is a VC configured for the ingress interface over which Layer 2 PDUs received from CE1 are forwarded, so it allocates a local VC label and binds it to VC ID configured under the ingress interface.
4. PE1 encodes this binding with the VC label TLV and VC FEC TLV and sends it to PE2 in a Label-Mapping message.
5. PE1 receives a Label-Mapping message from PE2 with a VC FEC TLV and VC label TLV. In the VC FEC TLV, the VC ID has a match with a locally configured VC ID. The VC label encoded in the VC label TLV is the outgoing VC label that PE1 is going to use when forwarding Layer 2 PDUs to PE2 for that particular VC.
6. PE1 might receive a Label-Request message from some other PE with a specific VC FEC TLV at any time during the OPERATIONAL state. PE1 examines the VC ID encoded in the FEC element, and responds to the peer a Label-Mapping with the local VC label corresponding to the VC ID.

7. PE2 performs the same Steps 1–6 as PE1. After both exchange the VC labels for a particular VC ID, the VC with that VC ID is fully established.
8. When one LSP of a VC is taken down for some reason, for example, the CE-PE link goes down or the VC configuration is removed from one PE router, the PE router must send a Label-Withdraw message to its remote LDP peer to withdraw the VC label it previously advertised.

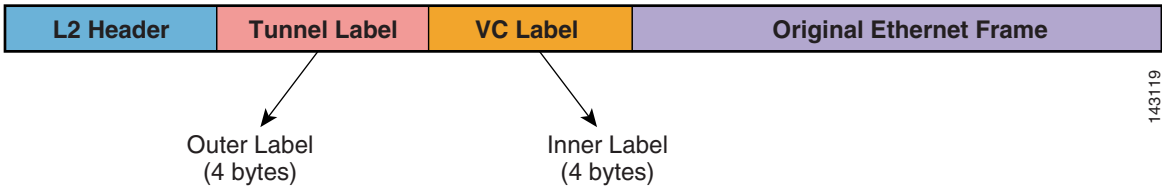
The VC is defined on the inner-most label and is used to bind to the interface to which the packet is to be delivered. The tunnel label is used to forward the packet through the network. Outer labels are assigned to this resultant frame at the egress PE interface for switching trough the MPLS core.

There are two VC types for EoMPLS: type 4 and type 5. Type 4 is used for extending VLANs over MPLS, while type 5 is useful for Ethernet port to port tunneling (port transparency).

Type 5 VCs are used for port-to-port transport of the entire frame without preamble or FCS. BPDUs are also carried transparently across the tunnel.

Only dot1Q vlan tagging is supported.

Figure 6-16 EoMPLS Header



Technical Requirements for EoMPLS

For the Cisco 7600 to perform as an LER to support transport of Layer 2 VLAN packets over MPLS, it must be able to forward 802.1QVLAN Layer 2 VCs across an IP/MPLS backbone. The features required to provide this capability are:

- Targeted LDP session between the peers
- Virtual Circuit between peers
- Two-level labeling
- Label imposition/disposition
- CoS mapping

LDP Session

The ingress PE needs a tunnel label from its upstream IGP neighbor to route the packets it receives from the ingress interface of the ingress PE to the egress PE. The ingress PE and its IGP neighbor are local label distribution peers and maintain a LDP session over which tunnel labels are distributed. They must all have the FEC that contains the host route for the egress PE and the tunnel label is bound to that FEC.

In addition, one targeted LDP session is required between the ingress PE and egress PE, which are remote label distribution peers, to exchange VC labels. All VC label bindings exchanged over this LDP session use the VC FEC element type 128 via the LDP “downstream unsolicited mode.” Because only a single LDP session is required, one is created only if not already present. A session may already be active because another application has already established a targeted LDP session between the two PE routers. LDP, which is only a transport protocol, shall determine that a LDP message is for the AToM application by checking for the existence of the VC FEC element type 128.

Support for the VC FEC element type 128 is required in the following LDP messages:

- Label Mapping Request
- Label Mapping
- Label Mapping Withdraw

If using conservative label retention, the ingress PE also needs to send Label-Request messages for all locally-configured VCs.

If using liberal label retention, the ingress PE does not need to send Label-Request messages for configured VCs, but must prepare to respond to Label-Request sent by other PEs which use conservative label retention. AToM is going to migrate to using liberal label retention mode, but initially the implementation employs conservative label retention mode.

Two Level Labeling

The Layer 2 transport service over MPLS is implemented through the use of two level label switching between the edge routers. The label used to route the packet over the MPLS backbone to the destination PE is called the "tunnel label." The label used to determine the egress interface is referred to as the VC label. The egress PE allocates a VC label and binds the Layer 2 egress interface to the VC in question, then it signals this label to the ingress PE via the targeted LDP session.

Label Imposition/Disposition

When the Layer 2 PDU arrives at the ingress interface of the ingress PE, the router must perform label imposition and switch the packet to the appropriate outgoing MPLS interface which routes the packet to the egress LER for the VC in question.

When the egress PE receives the packet, it receives the packet with only the VC label because its neighbor (known as the penultimate router) pops the tunnel label before forwarding the packet. The egress PE uses the VC label to perform disposition and switch the packet to the appropriate egress interface.

Class of Service (QoS) Mapping

MPLS provides QoS using the three experimental bits in a label to determine the queue of packets. To support QoS from PE to PE, the experimental bits in both the VC and Tunnel labels must be set. The experimental bits need to be set in the VC label because the tunnel label is popped at the penultimate router. In the case of EoMPLS, two methods of setting experimental bits are provided.

Static Setting of Experimental Bits

The customer is able to configure the PE to set the EXP bits in the labels to a given value based in ingress interface.

Using VLAN User Priority Bits to Determine Experimental Bit Settings

The three user priority bits are used to index into a table of eight values. The value for a given index is used to set the experimental bits. This method may cause out-of-order packets when packets have different user priorities.

Load Sharing—Ingress PE

Even there are multiple equal-cost routes between the two LERs, there is no load sharing done on ingress LER, because packets still have an Layer 2 header when EARL does the FIB lookup.

Load Sharing—MPLS Core

When EoMPLS packets get into the core of MPLS, the load sharing behavior may be different depending on how LSR looks into the packets. Assuming LSRs are also Constellation 2 systems, and then with control word inserted, EoMPLS packet flows are load shared among the equal paths because by checking the first nibble that is below the label stack, earl7 parses down as deep as the fifth or lowest label to use for hashing. Without the control word inserted, EoMPLS packet flows behavior over equal cost paths are different depends on the value of MAC address under the label stack. That is the reason why Constellation 2 and Cisco IOS inserts control word to EoMPLS packets by default.

Currently, the Cisco 7600 supports two main models of EoMPLS:

- PFC-based EoMPLS (No Local Switching)

  In this mode, the PFC3B/XL performs all label imposition /deposition functionality, as well as any required Ethernet encapsulation and 802.1Q header insertion.

  VC type is a result of negotiation between the two PEs, type 4 and 5 are both supported.

  IP ToS is always preserved end-end.

- PXF-based EoMPLS (With Local Switching)

  The egress interface is responsible for all label imposition and deposition, resulting in the requirement that this be an OSM interface on egress. The PFC simply forwards any packet received that contains a MAC address from the other side to the egress interface. The PFC does not take into consideration the EoMPLS overhead.

  For PXF-based EoMPLS, OSM cards are required for the core-facing interface.

  VC type is a result of negotiation between the two PEs; type 4 and 5 are both supported.

  This mode supports local switching, which may be useful for providing access-level redundancy.

  If ingress port is marked as Trust DSCP, CoS is not preserved.

  When using **mls qos queueing-only**, ingress PFC CoS and DSCP are preserved because you bypass the PFC QoS mechanisms; shaping and queuing are done on the egress OSM interface.

For CoS preservation, Type 5 VCs result in the stripping of the dot1Q header at the ingress PE, so there is no preservation. If at the egress PE, **mls qos** is enabled, the EXP bits are mapped back to the CoS bits. On ingress ports, use the command **mls qos trust cos**.

EoMPLS Restrictions

Dynamic IP Labeling

To support this feature, Dynamic IP labeling (**mpls ip**) must be enabled an all paths between the two imposition/disposition LERs. Failure to do so results in the packet being discarded before it reaches the disposition PE.

Summarized Routes

Routes from a PE discovered by its peers must be unsummarized; that is, address/32. This is required to ensure that there is an LSP from PE to PE.

Frame Size

The MPLS network should be configured with an MTU that is at least of 12 bytes plus link header larger than the largest frame size that is transported in the LSPs.

CE-side MTU or SP core links must be changed to accommodate the encapsulation overhead.

If a packet length, after it has been encapsulated on the ingress LSR, exceeds the LSP MTU, it must be dropped.

If an egress LSR receives a packet on an EoMPLS VC with a length, after the label stack and control word have been popped, that exceeds the MTU of the egress Layer 2 interface (VLAN), it must be dropped.

Fragmentation is not supported for Layer 2 packets transmitted across the MPLS backbone. Therefore to deploy the layer transport service, the network operator must make sure that the MTU of all intermediate links between the endpoints is sufficient to carry the largest Layer 2 transported packets that are received. The MTU setting of the ingress PE needs to match with the setting at the egress PE. A VC does not properly establish if MTU sizes mismatch.

Configuration and Monitoring

For this design guide, all testing on the Cisco 7600 was done with PFC-based EoMPLs, rather than PXF-based.

PXF-Based Cisco 7600 Configuration

```
class-map match-any L2VPN_TRAFFIC
  match any
policy-map SET_L2VPN_TRAFFIC
  class L2VPN_TRAFFIC
   set mpls experimental 3
interface Loopback0
ip address 1.1.1.1 255.255.255.255

interface GigabitEthernet9/1.1
encapsulation dot1Q 2000
xconnect 2.2.2.2 200 encapsulation mpls
service-policy input SET_L2VPN_TRAFFIC
interface GE-WAN9/3 (core facing interface)
ip address 10.10.93.2 255.255.255.0
negotiation auto
tag-switching ip
mls qos trust dscp
```

Cisco 12K Configuration

```
interface Loopback0
ip address 1.1.1.1 255.255.255.255   !Loopback address Must be /32
interface GigabitEthernet9/1
mtu 9216
interface GigabitEthernet9/1.1
encapsulation dot1Q 2000 this can be used
xconnect 2.2.2.2 200 encapsulation mpls    ! Remote Loopback addr and VC id
```

Cisco 7200 Configuration

```
class-map match-any L2VPN_TRAFFIC
  match any
policy-map SET_L2VPN_TRAFFIC
  class L2VPN_TRAFFIC
   set mpls experimental 3
```

```
interface Loopback0
ip address 10.191.44.251 255.255.255.255
interface GigabitEthernet0/2
 mtu 9216
 no ip address
 load-interval 30
 duplex full
 speed 1000
 media-type rj45
 no negotiation auto
!
interface GigabitEthernet0/2.20
 encapsulation dot1Q 20
 no snmp trap link-status
 no cdp enable
 xconnect 10.191.44.252 20  encapsulation mpls
 service-policy input SET_L2VPN_TRAFFIC
```

Cisco 3750 Metro Configuration

```
class-map match-any L2VPN_TRAFFIC
  match any
policy-map SET_L2VPN_TRAFFIC
  class L2VPN_TRAFFIC
   set mpls experimental 3

interface Loopback0
ip address 10.191.44.252 255.255.255.255
interface GigabitEthernet0/2
 mtu 9216
!
interface GigabitEthernet0/2.20
 encapsulation dot1Q 20
 no snmp trap link-status
 no cdp enable
 xconnect 10.191.44.251 20  encapsulation mpls
 service-policy input SET_L2VPN_TRAFFIC
```

Cisco PXF-Based and Cisco 12K Monitoring Commands

```
7600#sh mpls l2transport vc
Local intf Local circuit Dest address VC ID Status
------------- -------------------- --------------- ---------- ----------
Gi9/1.1 Eth VLAN 2000 2.2.2.2 200 UP VC Status

7600#sh mpls l2transport vc detail
Local interface: Gi9/1.1 up, line protocol up, Eth VLAN 2000 up
Destination address: 2.2.2.2, VC ID: 200, VC status: up
Tunnel label: 16, next hop 10.10.93.1
Output interface: GE9/3, imposed label stack {16 2000} Two level label
Create time: 01:37:35, last status change time: 00:00:13
Signaling protocol: LDP, peer 2.2.2.2:0 up
MPLS VC labels: local 21, remote 2000
Group ID: local 0, remote 0
MTU: local 1500, remote 1500
Remote interface description:
Sequencing: receive disabled, send disabled
VC statistics:
packet totals: receive 234343, send 53196336
```

```
byte totals: receive 232323332, send 3191872568
packet drops: receive 0, send 0

7600#remote command switch show mpls l2transport vc detail
Local interface: GigabitEthernet9/1, Eth VLAN 2000
Destination address: 2.2.2.2, VC ID: 200 VC TYPE
VC status: receive UP, send UP
VC type: receive 5, send 5
Tunnel label: 16, next hop 10.10.93.1
Output interface: GE9/3, imposed label stack {16 2000}
MPLS VC label: local 21, remote 2000
Linecard VC statistics:
packet totals: receive: 234343 send: 38222872
byte totals: receive: 232323332 send: 2293372320
packet drops: receive: 0 send: 0

7600#sh mpls l2transport binding
Destination Address: 2.2.2.2, VC ID: 200
Local Label: 21
Cbit: 0, VC Type: Ethernet, GroupID: 0
MTU: 1500, Interface Desc: n/a
Remote Label: 2000
Cbit: 0, VC Type: Ethernet, GroupID: 0
MTU: 1500, Interface Desc: n/a

7600#sh mls cef eom
Index VPN Adjacency
128 257 278528,0
7600#sh mls cef adjacency entry 278528 detail
Index: 278528 smac: 000b.fcd4.cf00, dmac: 00d0.0362.7800
mtu: 1518, vlan: 1020, dindex: 0x0, l3rw_vld: 1
format: MPLS, flags: 0xD000008400 Hardware Programmed MTU
label0: 0, exp: 0, ovr: 0
label1: 2000, exp: 0, ovr: 0
label2: 16, exp: 0, ovr: 0
op: PUSH_LABEL2_LABEL1
packets: 69787014, bytes: 4187220840
```

Cisco PFC-Based Configuration

```
vlan 2000  !configure vlan in global database
interface GigabitEthernet9/1 ! CE facing port as Access/Trunk Port
no ip address
switchport
switchport trunk encapsulation dot1q
switchport mode trunk
interface Vlan2000
no ip address
xconnect 1.1.1.1 200 encapsulation mpls
```

Cisco PXF-Based Monitoring Commands

```
7600#sh mpls l2transport vc
Local intf Local circuit Dest address VC ID Status
------------ -------------------- --------------- ---------- ----------
Vl500 Eth VLAN 500 1.1.1.1 200 UP

7600#sh interfaces trunk
Port Mode Encapsulation Status Native vlan
Gi8/2 on 802.1q trunking 1
Port Vlans allowed on trunk
```

```
Gi8/2 1-4094
Port Vlans allowed and active in management domain
Gi8/2 1,2000
Port Vlans in spanning tree forwarding state and not pruned
Gi8/2 1,2000
7600#ipc-con 8 0    ⇓Remote Login to PXF card Imposing label
Entering CONSOLE for slot 8
Type "^C^C^C" to end this session
CWTLC-Slot8>en
CWTLC-Slot8#sh mpls l2transport vc
Local intf Local circuit Dest address VC ID Receive/Send
VC Status
------------- ------------------ --------------- ---------- ------------
Vl2000 Eth VLAN 500 1.1.1.1 200 UP /UP
CWTLC-Slot8#sh mpls l2transport vc de
Local interface: Vlan2000, Eth VLAN 2000
Destination address: 1.1.1.1, VC ID: 200
VC status: receive UP, send UP
VC type: receive 5, send 5
Tunnel label: 16, next hop 10.10.191.2
Output interface: GE8/1, imposed label stack {16 23}
MPLS VC label: local 21, remote 23
Linecard VC statistics:
packet totals: receive: 0 send: 0
byte totals: receive: 0 send: 0
packet drops: receive: 0 send: 0
Control flags:
receive 1, send: 11
CWTLC EoMPLS disp detailed info: AC if_no 30
t vclbl VLAN Type h impidx stat
- d------- x---(d---) ------- - x-- x---
0 00000021 07D0(2000) ether 1 1C4 0001
1 00000021 07D0(2000) ether 1 1C4 0001
vlan(2000) rx_pkts(0)
CWTLC EoMPLS imp detailed info: AC if_no 30, Egress GE-WAN8/1
Vlan func[0]: 2000(0x7D0): flag(0x0) func(3:atom ether) hash (0x1)
Tx TVC Table:
idx ltl op vcinfo en h next intfid
x--- x-- -- d----- -- - x--- x-----
tx-tvc 0044 004 02 000021 00 1 0000 000000 pxf[0] vlan: 2000 hash:001
TTFIB: Index(21) Imposition(PUSH 2):{16, 23, 0, 0}    ⇓PXF Entries
t VLAN vc lbl tun lbl MAC Address cos2exp
```