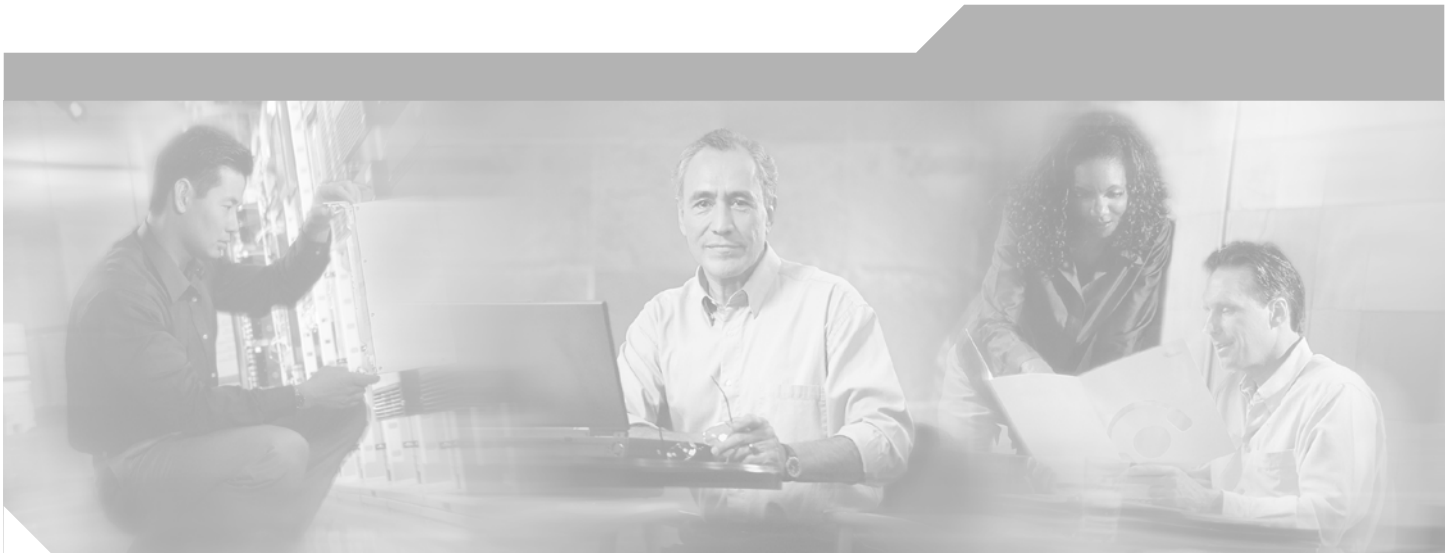


Cisco IP Telephony Solution Reference Network Design

Cisco CallManager Release 3.3
November 2003

Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

Customer Order Number: 956662

Preface

This document provides design considerations and guidelines for implementing Cisco IP Telephony solutions based on the Cisco Architecture for Voice, Video, and Integrated Data (AVVID).

This document is primarily an update of the design guidelines and information presented in the *Cisco IP Telephony Solution Reference Network Design* (SRND) for Cisco CallManager releases 3.1 and 3.2, which is available online at

http://cisco.com/go/srnd

This document assumes that you are already familiar with the terms and concepts presented in previous versions of the *Cisco IP Telephony SRND*. If you want to review any of those terms and concepts, refer to the documentation at the preceding URL.

New or Changed Information for This Release

Unless stated otherwise, the information in this document applies specifically to Cisco CallManager Release 3.3. Table 1 lists the features and design considerations that are new for this release or that have changed significantly from previous releases of Cisco CallManager.

Table 1 New or Changed Information for Cisco CallManager Release 3.3

Topic	Described in:
Accessibility and Section 508 conformance	Design Considerations for Section 508 Conformance, page 1-28
Alternate gatekeeper (gatekeeper clustering)	Gatekeeper Considerations, page 6-10 Gatekeeper Clustering (Alternate-Gatekeeper), page 6-21
Automated alternate routing (AAR)	Automated Alternate Routing, page 7-9
Call processing	Call Processing with Cisco CallManager Release 3.3, page 6-2
Call processing redundancy	Call Processing Redundancy, page 6-7
Calling search spaces	An Alternative Approach to Configuring Calling Search Spaces, page 7-8
Dial plan weights	Dial Plan Weights, page 6-5
Emergency services (911)	Emergency Services, page 8-1
Extension mobility	Special Considerations for Extension Mobility, page 7-9
Fax and modem support	Fax and Modem Support, page 3-12
Hardware and software recommendations	Recommended Hardware and Software Combinations, page A-1

Table 1 New or Changed Information for Cisco CallManager Release 3.3 (continued)

Topic	Described in:
High performance server	Clustering Guidelines, page 6-1 Server Platforms, page 6-4
Intercluster gatekeeper-controlled trunk with Cisco CallManager locations	Intercluster Gatekeeper-Controlled Trunk with Locations, page 1-15
Media resources	Media Resources, page 4-1
Multiprotocol Label Switching (MPLS)	Multi-Site MPLS WAN Considerations, page 1-20 WAN Infrastructure, page 2-4
Music on hold	Music on Hold, page 5-1
QSIG	QSIG Support, page 3-11
Security considerations	Security, page 15-1
Trivial File Transfer Protocol (TFTP)	Multi-Cluster Campus TFTP Services, page 1-25
Voice over the PSTN (VoPSTN)	Voice Over the PSTN as a Variant of Centralized Call Processing, page 1-7

Revision History

The following table lists the revision history for this document.

Revision Date	Comments
November, 2003	The following sections are new or have been updated since the previous release of this document: • Voice Over the PSTN as a Variant of Centralized Call Processing, page 1-7 • Multi-Cluster Campus TFTP Services, page 1-25 • Music on Hold, page 5-1 • Automated Alternate Routing, page 7-9 • Emergency Services, page 8-1
September, 2003	Revisions for Cisco CallManager Release 3.3(3). The following sections are new or have been updated since the previous release of this document: • Multi-Site MPLS WAN Considerations, page 1-20 • Design Considerations for Section 508 Conformance, page 1-28 • Fax and Modem Support, page 3-12 • Media Resources, page 4-1 • Music on Hold, page 5-1 • Emergency Services, page 8-1 • Security, page 15-1
April, 2003	Initial draft.

Obtaining Documentation

Cisco provides several ways to obtain documentation, technical assistance, and other technical resources. These sections explain how to obtain technical information from Cisco Systems.

Cisco.com

You can access the most current Cisco documentation on the World Wide Web at this URL:

http://www.cisco.com/univercd/home/home.htm

You can access the Cisco website at this URL:

http://www.cisco.com

International Cisco web sites can be accessed from this URL:

http://www.cisco.com/public/countries_languages.shtml

Documentation CD-ROM

Cisco documentation and additional literature are available in a Cisco Documentation CD-ROM package, which may have shipped with your product. The Documentation CD-ROM is updated monthly and may be more current than printed documentation. The CD-ROM package is available as a single unit or through an annual subscription.

Registered Cisco.com users can order the Documentation CD-ROM (product number DOC-CONDOCCD=) through the online Subscription Store:

http://www.cisco.com/go/subscription

Ordering Documentation

You can find instructions for ordering documentation at this URL:

http://www.cisco.com/univercd/cc/td/doc/es_inpck/pdi.htm

You can order Cisco documentation in these ways:

- Registered Cisco.com users (Cisco direct customers) can order Cisco product documentation from the Networking Products MarketPlace:

  http://www.cisco.com/en/US/partner/ordering/index.shtml

- Registered Cisco.com users can order the Documentation CD-ROM (Customer Order Number DOC-CONDOCCD=) through the online Subscription Store:

  http://www.cisco.com/go/subscription

- Nonregistered Cisco.com users can order documentation through a local account representative by calling Cisco Systems Corporate Headquarters (California, U.S.A.) at 408 526-7208 or, elsewhere in North America, by calling 800 553-NETS (6387).

Documentation Feedback

You can submit comments electronically on Cisco.com. On the Cisco Documentation home page, click **Feedback** at the top of the page.

You can e-mail your comments to bug-doc@cisco.com.

You can submit your comments by mail by using the response card behind the front cover of your document or by writing to the following address:

Cisco Systems
Attn: Customer Document Ordering
170 West Tasman Drive
San Jose, CA 95134-9883

We appreciate your comments.

Obtaining Technical Assistance

Cisco provides Cisco.com, which includes the Cisco Technical Assistance Center (TAC) Website, as a starting point for all technical assistance. Customers and partners can obtain online documentation, troubleshooting tips, and sample configurations from the Cisco TAC website. Cisco.com registered users have complete access to the technical support resources on the Cisco TAC website, including TAC tools and utilities.

Cisco.com

Cisco.com offers a suite of interactive, networked services that let you access Cisco information, networking solutions, services, programs, and resources at any time, from anywhere in the world.

Cisco.com provides a broad range of features and services to help you with these tasks:

- Streamline business processes and improve productivity
- Resolve technical issues with online support
- Download and test software packages
- Order Cisco learning materials and merchandise
- Register for online skill assessment, training, and certification programs

To obtain customized information and service, you can self-register on Cisco.com at this URL:

http://www.cisco.com

Technical Assistance Center

The Cisco TAC is available to all customers who need technical assistance with a Cisco product, technology, or solution. Two levels of support are available: the Cisco TAC website and the Cisco TAC Escalation Center. The avenue of support that you choose depends on the priority of the problem and the conditions stated in service contracts, when applicable.

We categorize Cisco TAC inquiries according to urgency:

- Priority level 4 (P4)—You need information or assistance concerning Cisco product capabilities, product installation, or basic product configuration.
- Priority level 3 (P3)—Your network performance is degraded. Network functionality is noticeably impaired, but most business operations continue.
- Priority level 2 (P2)—Your production network is severely degraded, affecting significant aspects of business operations. No workaround is available.
- Priority level 1 (P1)—Your production network is down, and a critical impact to business operations will occur if service is not restored quickly. No workaround is available.

Cisco TAC Website

You can use the Cisco TAC website to resolve P3 and P4 issues yourself, saving both cost and time. The site provides around-the-clock access to online tools, knowledge bases, and software. To access the Cisco TAC website, go to this URL:

http://www.cisco.com/tac

All customers, partners, and resellers who have a valid Cisco service contract have complete access to the technical support resources on the Cisco TAC website. Some services on the Cisco TAC website require a Cisco.com login ID and password. If you have a valid service contract but do not have a login ID or password, go to this URL to register:

http://tools.cisco.com/RPF/register/register.do

If you are a Cisco.com registered user, and you cannot resolve your technical issues by using the Cisco TAC website, you can open a case online at this URL:

http://www.cisco.com/en/US/support/index.html

If you have Internet access, we recommend that you open P3 and P4 cases through the Cisco TAC website so that you can describe the situation in your own words and attach any necessary files.

Cisco TAC Escalation Center

The Cisco TAC Escalation Center addresses priority level 1 or priority level 2 issues. These classifications are assigned when severe network degradation significantly impacts business operations. When you contact the TAC Escalation Center with a P1 or P2 problem, a Cisco TAC engineer automatically opens a case.

To obtain a directory of toll-free Cisco TAC telephone numbers for your country, go to this URL:

http://www.cisco.com/warp/public/687/Directory/DirTAC.shtml

Before calling, please check with your network operations center to determine the level of Cisco support services to which your company is entitled: for example, SMARTnet, SMARTnet Onsite, or Network Supported Accounts (NSA). When you call the center, please have available your service agreement number and your product serial number.

Obtaining Additional Publications and Information

Information about Cisco products, technologies, and network solutions is available from various online and printed sources.

- *The Cisco Product Catalog* describes the networking products offered by Cisco Systems as well as ordering and customer support services. Access the *Cisco Product Catalog* at this URL:

  http://www.cisco.com/en/US/products/products_catalog_links_launch.html

- Cisco Press publishes a wide range of networking publications. Cisco suggests these titles for new and experienced users: *Internetworking Terms and Acronyms Dictionary, Internetworking Technology Handbook, Internetworking Troubleshooting Guide, and the Internetworking Design Guide*. For current Cisco Press titles and other information, go to Cisco Press online at this URL:

  http://www.ciscopress.com

- *Packet* magazine is the Cisco monthly periodical that provides industry professionals with the latest information about the field of networking. You can access *Packet* magazine at this URL:

  http://www.cisco.com/en/US/about/ac123/ac114/about_cisco_packet_magazine.html

- *iQ Magazine* is the Cisco monthly periodical that provides business leaders and decision makers with the latest information about the networking industry. You can access *iQ Magazine* at this URL:

  http://business.cisco.com/prod/tree.taf%3fasset_id=44699&public_view=true&kbns=1.html

- *Internet Protocol Journal* is a quarterly journal published by Cisco Systems for engineering professionals involved in the design, development, and operation of public and private internets and intranets. You can access the *Internet Protocol Journal* at this URL:

  http://www.cisco.com/en/US/about/ac123/ac147/about_cisco_the_internet_protocol_journal.html

- Training—Cisco offers world-class networking training, with current offerings in network training listed at this URL:

  http://www.cisco.com/en/US/learning/le31/learning_recommended_training_list.html

IP Telephony Deployment Models

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

- Single Site, page 1-2

  The single-site model for IP telephony consists of a call processing agent located at a single site and a LAN or metropolitan area network (MAN) to carry voice traffic throughout the site. Calls beyond the LAN or MAN use the public switched telephone network (PSTN). If an IP WAN is incorporated into the single-site model, it is for data traffic only; no telephony services are provided over the WAN.

  Use this model for a single campus or site with less than 30,000 lines.

- Multi-Site WAN with Centralized Call Processing, page 1-4

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

  Use this model for a main site with many smaller remote sites that are connected via a QoS-enabled WAN but that do not require full features and functionality during a WAN outage.

- Multi-Site WAN with Distributed Call Processing, page 1-9

  The multi-site WAN model with distributed call processing consists of multiple independent sites, each with its own call processing agent connected to an IP WAN that carries voice traffic between the distributed sites. The IP WAN in this model does not carry call control signaling between the sites because each site has its own call processing agent.

  Use this model for a large central site with more than 30,000 lines or for a deployment with more than six large sites (more than 30,000 lines total) interconnected via a QoS-enabled WAN.

- Clustering Over the IP WAN, page 1-17

  This model deploys a single Cisco CallManager cluster across multiple sites that are connected by an IP WAN with QoS features enabled.

  Use this model for a deployment with a maximum of six large sites (maximum of 30,000 lines total) interconnected via a QoS-enabled WAN.

Note Other sections of this document assume that you understand the concepts involved with these deployment models, so please become thoroughly familiar with them before proceeding.

In addition, this chapter describes the following special design considerations and variations to the main deployment models:

- Multi-Site MPLS WAN Considerations, page 1-20

  This section describes how to adapt the IP Telephony deployment models to support a full-mesh routing technology such as Cisco IOS Multiprotocol Label Switching (MPLS).

- Multi-Cluster Campus TFTP Services, page 1-25

  This section describes how to use a single TFTP server to service multiple clusters and how to distribute TFTP functionality across multiple servers to provide load balancing and redundancy.

- Design Considerations for Section 508 Conformance, page 1-28

  This section presents guidelines for designing you IP telephony network to provide accessibility to users with disabilities, in conformance with U.S. Section 508.

Single Site

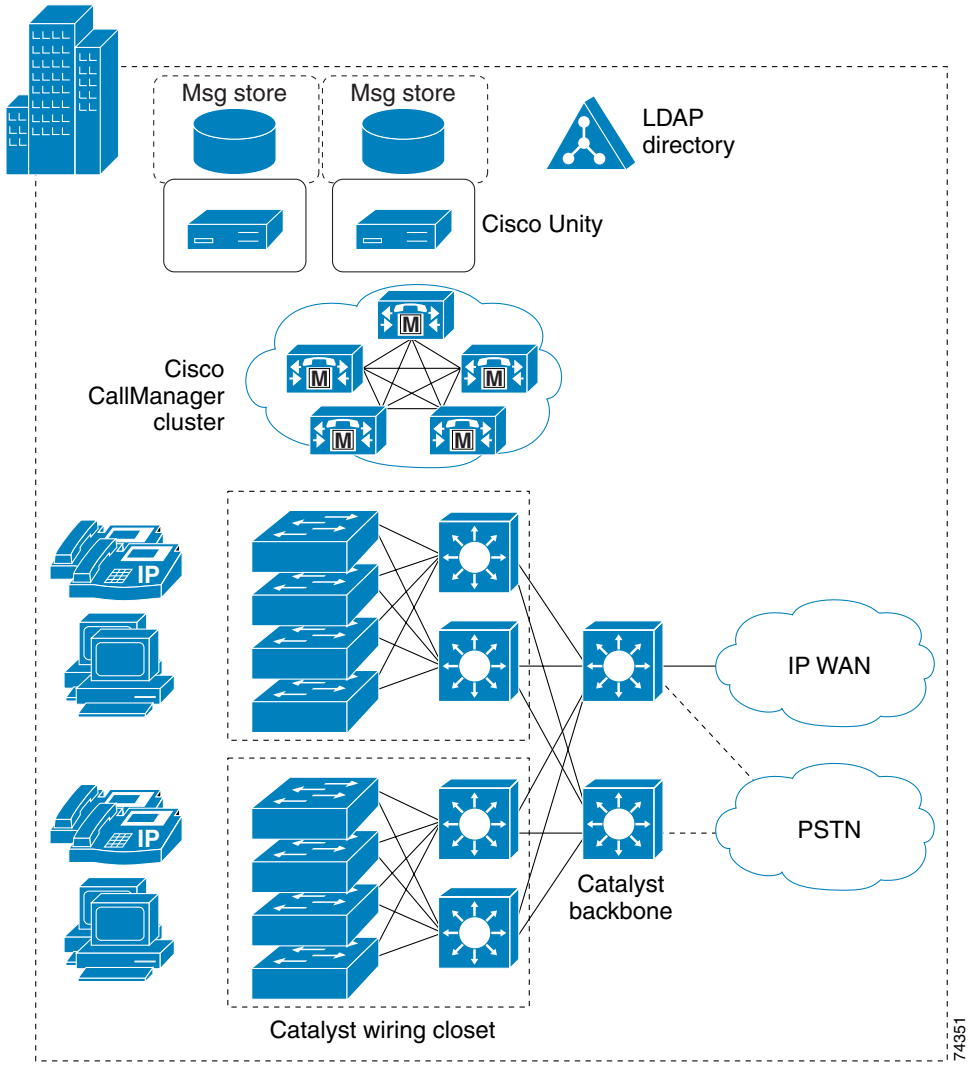
The single-site model for IP telephony consists of a call processing agent located at a single site, or campus, with *no* telephony services provided over an IP WAN. An enterprise would typically deploy the single-site model over a LAN or metropolitan area network (MAN), which carries the Voice over IP (VoIP) traffic within the site. In this model, calls beyond the LAN or MAN use the public switched telephone network (PSTN).

The single-site model has the following design characteristics:

- Single Cisco CallManager or Cisco CallManager cluster
- Maximum of 30,000 IP phones per cluster
- PSTN for all external calls
- Digital signal processor (DSP) resources for conferencing, transcoding, and media termination point (MTP)
- Voice mail and unified messaging components
- Only G.711 codecs for all IP phone calls (80 kbps of IP bandwidth per call, uncompressed)
- Capability to integrate with legacy private branch exchange (PBX) and voice mail systems

Figure 1-1 illustrates the model for an IP telephony network within a single campus or site.

Figure 1-1 Single-Site Model



Best Practices for the Single-Site Model

Follow these guidelines and best practices when implementing the single-site 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, and expansion of your IP telephony deployment across the WAN or to multiple Cisco CallManager clusters.
- Know the calling patterns for your enterprise. Use the single-site model if most of the calls from your enterprise are within the same site or to PSTN users outside your enterprise.
- Use G.711 codecs for all endpoints. This practice eliminates the consumption of digital signal processor (DSP) resources for transcoding, and those resources can be allocated to other functions such as conferencing and Media Termination Points (MTPs).

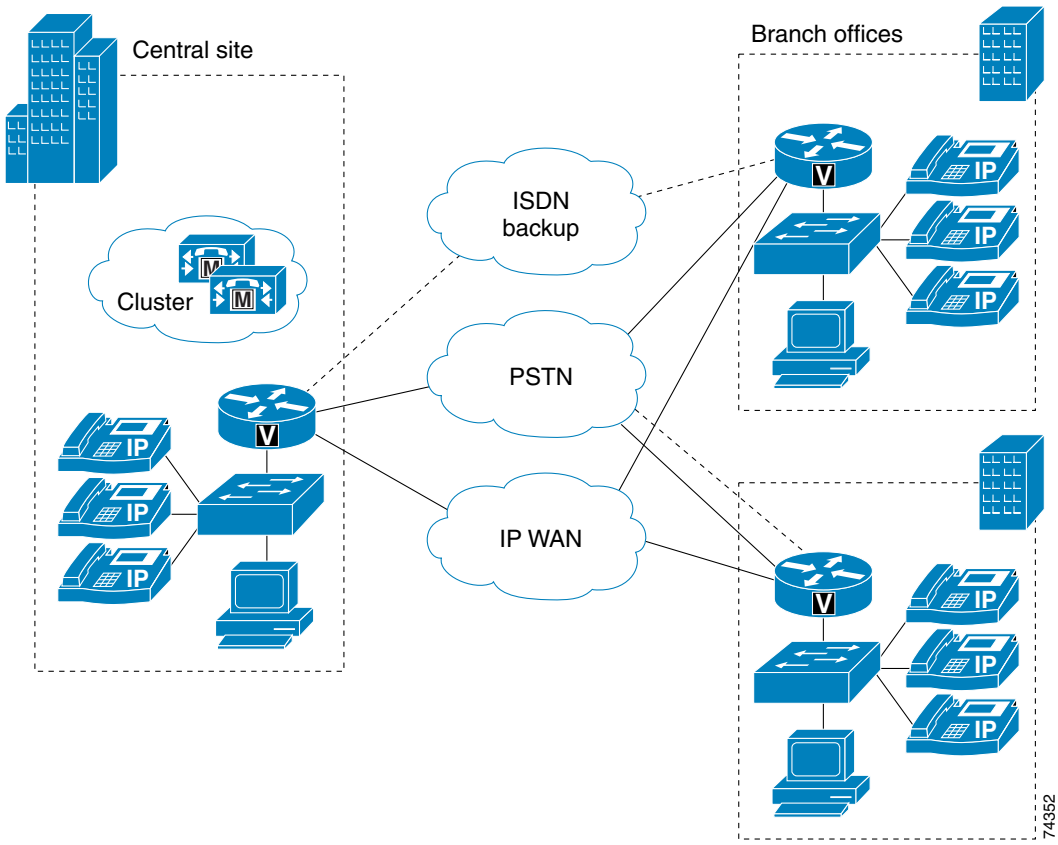
- Use Media Gateway Control Protocol (MGCP) gateways for the PSTN if you do *not* require H.323 functionality. This practice simplifies the dial plan configuration. H.323 might be required to support specific functionality not offered with MGCP, such as support for Signaling System 7 (SS7) or Non-Facility 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 Network Infrastructure, page 2-1.)
- Follow the provisioning recommendations listed in the section on Call Processing, page 6-1.

Multi-Site WAN with Centralized Call Processing

The multi-site WAN model with centralized call processing consists of a single call processing agent that provides services for many sites and uses the IP WAN to transport IP telephony traffic between the sites. The IP WAN also carries call control signaling between the central site and the remote sites. Figure 1-2 illustrates a typical centralized call processing deployment, with a Cisco CallManager cluster as the call processing agent at the central site and an IP WAN with QoS enabled to connect all the sites. The remote sites rely on the centralized Cisco CallManager cluster to handle their call processing. Applications such as voice mail and Interactive Voice Response (IVR) systems 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 an IP WAN with QoS enabled.

Figure 1-2 Centralized Call Processing Deployment Model



Connectivity options for the IP WAN include:

- Leased lines
- Frame Relay
- Asynchronous Transfer Mode (ATM)
- ATM and Frame Relay Service Inter-Working (SIW)
- Multiprotocol Label Switching (MPLS) Virtual Private Network (VPN)
- Voice and Video Enabled IP Security Protocol (IPSec) VPN (V3PN)

Routers that reside at the WAN edges require quality of service (QoS) mechanisms, such as priority queuing and traffic shaping, to protect the voice traffic from the data traffic across the WAN, where bandwidth is typically scarce. In addition, a call admission control scheme is needed to avoid oversubscribing the WAN 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. (Refer to the section on Call Admission Control for Centralized Call Processing, page 1-6, for more information on locations.)

A variety of Cisco gateways can provide the remote sites with PSTN access. When the IP WAN is down, or if all the available bandwidth on the IP WAN has been consumed, users at the remote sites can dial the PSTN access code and place their calls through the PSTN. The Survivable Remote Site Telephony (SRST) feature, available on Cisco IOS gateways, provides call processing at the branch offices in the event of a WAN failure.

Note It is possible to use other WAN technologies that lack some of the QoS features required for converged voice and data traffic, but these technologies have special design considerations that are beyond the scope of this document. In addition, those other technologies usually do not maintain good voice quality due to their lack of QoS features.

Best Practices for the Multi-Site Model with Centralized Call Processing

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

- Minimize delay between Cisco CallManager and remote locations to reduce voice cut-through delays (also known as clipping).
- For hub-and-spoke topologies, use the locations mechanism in Cisco CallManager for call admission control into and out of remote branches. If the WAN uses Cisco IOS Multiprotocol Label Switching (MPLS), see the section on Multi-Site MPLS WAN Considerations, page 1-20.
- 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 Survivable Remote Site Telephony (SRST) mode at each remote site depends on the branch router platform, the amount of memory installed, and the Cisco IOS release. (For the latest SRST platform and code specifications, refer to the SRST documentation at 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:
  - IP WAN bandwidth or delay limitations
  - Criticality of the voice network
  - Feature set needs
  - Scalability
  - Ease of management
  - Cost

  If a distributed call processing model is deemed more suitable for the customer's business needs, the choices include installing a local Cisco CallManager server or running the Cisco IOS Telephony Service (ITS) on the branch router.

Call Admission Control for Centralized Call Processing

Multi-site deployments require some form of call admission control to ensure the voice quality of calls transmitted across network links that have limited available bandwidth. Cisco CallManager provides a simple mechanism know as *locations* for implementing call admission control in multi-site WAN deployments with centralized call processing. Follow these guidelines when using locations for call admission control:

- Locations require a hub-and-spoke network topology.
- Configure a separate location in Cisco CallManager for each site.

- Configure the appropriate bandwidth limit for each site according to the type of codec used at that site. (See Table 1-1 for bandwidth settings.)
- Assign each device configured in Cisco CallManager to a location. If you move a device to another location, change its location configuration as well.
- Cisco CallManager supports up to 500 locations.

Prior to Cisco CallManager Release 3.1, a cluster could support only one primary (active) Cisco CallManager server when using locations for call admission control. With Cisco CallManager Release 3.1 and later, the locations bandwidth is shared among all Cisco CallManager subscriber servers in the cluster, thus enabling you to use the locations mechanism with any size cluster.

Table 1-1 Bandwidth Settings by Codec Type

	Codec Type	
Parameter Setting	**G.729**	**G.711**
Codec bit rate	8 kbps	64 kbps
Cisco CallManager locations	24 kbps	80 kbps
Cisco CallManager gatekeeper	16 kbps	128 kbps
Cisco IOS gateways, prior to release 12.2(2)XA	64 kbps	64 kbps
Cisco IOS gateways, release 12.2(2)XA and later	16 kbps	128 kbps

Voice Over the PSTN as a Variant of Centralized Call Processing

The centralized call processing deployment model can be adapted so that inter-site voice media is sent over the PSTN instead of the WAN. With this configuration, the signaling (call control) of all telephony endpoints is still controlled by the central Cisco CallManager cluster, therefore voice over the PSTN (VoPSTN) still requires a QoS-enabled WAN with appropriate bandwidth configured for the signaling traffic. VoPSTN also requires the use of the automated alternate routing (AAR) feature. (For more information on AAR, see the section on Automated Alternate Routing, page 7-9.)

To use the PSTN as the primary (and only) voice path, you can configure the call admission control bandwidth of each location (branch site) to 1 kbps, thus preventing *all* calls from traversing the WAN. With this configuration, all inter-site calls trigger the AAR functionality, which routes the calls over the PSTN.

VoPSTN offers basic voice functionality that is a reduced subset of the Cisco CallManager feature set.

Note In some instances, VoPSTN might not support all of the features normally afforded by the centralized call processing deployment model.

When considering a VoPSTN deployment, the system designer should address the following issues, among others:

- AAR functionality must be configured properly.
- As a general rule, supported call initiation endpoints include IP phones, gateways, and line-side gateway-driven analog phones.
- Inter-branch calls can use AAR only if the destination endpoints are IP phones or Cisco Unity ports. Inter-branch calls to other endpoints must use a fully qualified E.164 number.
- Centralized voice mail and unified messaging require:
  - A telephony network provider that supports redirected dialed number identification service (RDNIS) end-to-end for all locations that are part of the deployment. RDNIS is required so that calls redirected to voice mail carry the redirecting DN, to ensure proper voice mail box selection.
  - If the voice mail system is accessed through an MGCP gateway, the voice mail pilot number must be a fully qualified E.164 number.
- VoPSTN does not support the Extension Mobility feature.
- All on-net (intra-cluster) calls will be delivered to the destination phone with the same call treatment as an off-net (PSTN) call. This includes the quantity of digits delivered in the call directories such as Missed Calls and Received Calls.
- Each inter-branch call generates two independent call detail records (CDRs): one for the call leg from the calling phone to the PSTN and the other for the call leg from the PSTN to the called phone.
- There is no way to distinguish the ring type for on-net and off-net calls.
- All on-net, inter-branch calls will display the message, "Network congestion, rerouting."
- Do not implement shared lines across branches.
- Within a single branch, shared lines should be implemented as part of a partition reachable by the calling search spaces of devices (including the branch's PSTN gateway) within the same branch only. The home partition of the shared line DN should not be part of a calling search space of any other branch. Inter-branch access to the shared line DN should be through a translation pattern to a fully qualified PSTN number.
- All destination phones require a fully qualified Direct Inward Dial (DID) PSTN number that can be called directly. Non-DID DNs cannot be reached directly.
- If destination phones become unregistered (for example, due to WAN connectivity interruption), AAR functionality will *not* be invoked. If the destination phone has access to an SRST router, then it can be reached by directly dialing its PSTN DID number.
- With VoPSTN, music on hold (MoH) is limited to cases where the holding party is co-located with the MoH resource. If MoH is deployed at the central site, then only calls held by devices at the central site will receive the hold music.
- Transfers to a destination outside the branch site will result in the hairpinning of the call through the branch's gateway. Traffic engineering of the branch's gateway resources must be adjusted accordingly.
- Call forwarding of any call to a destination outside the branch site will result in the hairpinning of the call through the branch's gateway. This behavior includes calls forwarded to a voice mail system located outside the branch.
- Conferencing resources must be co-located with the conferencing phone because branch office phones will not have access to centralized DSP resources.

- VoPSTN does not support applications that require streaming of IP audio from the central site (that is, not traversing a gateway). These applications include, but are not limited to:
  - Centralized music on hold (MOH) servers
  - Interactive Voice Response (IVR)
  - CTI-based applications
- Cisco recommends that you do *not* use the Attendant Console outside of the central site because it requires a considerable amount of bandwidth to allow NT user account access into the WAUsers directory on Cisco CallManager.
- Because all inter-branch media (including transfers) is sent through the PSTN, the gateway trunk group must be sized to accommodate all inter-branch traffic, transfers, and centralized voice mail access.
- Cisco recommends that you do *not* deploy shared lines across branches, such that the devices sharing the line are in different branches.
- Shared lines within the same branch should be configured in a partition included only in that branch's calling search spaces. Inter-site access to the shared line requires one of the following:
  - The originating site dials the DID number of the shared line.
  - If inter-site abbreviated dialing to the shared line is desired, use a translation pattern that expands the user-dialed abbreviated string to the DID number of the shared line.

  Note In this case, direct dialing of the shared line's DN from another branch would trigger multiple AAR-based PSTN calls.
- Call Forward All functionality results in hairpinned calls through the local branch gateway in either one of the following cases:
  - Calls are forwarded to an external PSTN number.
  - Calls are forwarded to an on-net abbreviated dialing destination located in a different branch. In this case, Cisco recommends requiring the user to enter the fully qualified PSTN number of the destination.

Multi-Site WAN with Distributed Call Processing

The multi-site WAN model with distributed call processing consists of multiple independent sites, each with its own call processing agent connected to an IP WAN that carries voice traffic between the distributed sites. Unlike the centralized call processing model, however, the IP WAN in the distributed model does not carry call control signaling between the sites because each site has its own call processing agent. Figure 1-3 illustrates a typical distributed call processing deployment.

Each site in the distributed call processing model can be one of the following:

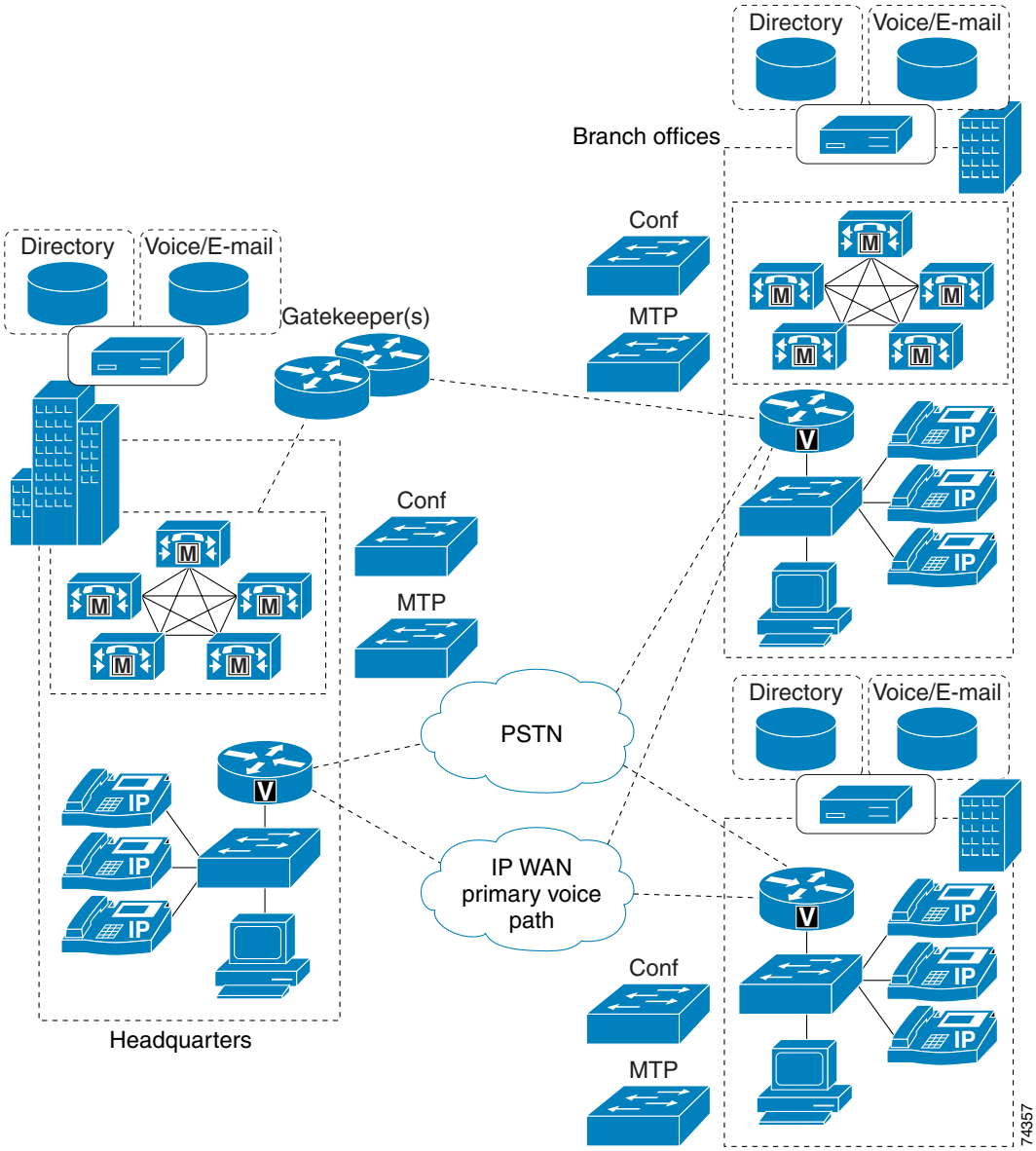
- A single site with its own call processing agent, which can be either Cisco CallManager, Cisco IOS Telephony Services (ITS), or other IP PBX
- A centralized call processing site and all of its associated remote sites
- A legacy PBX with Voice over IP (VoIP) gateway

An IP WAN interconnects all the distributed call processing sites. Typically, the PSTN serves as a backup connection between the sites in case the IP WAN connection fails or does not have any more available bandwidth. A site connected only through the PSTN is a standalone site and is not covered by the distributed call processing model. (See Single Site, page 1-2.)

Connectivity options for the IP WAN include:

- Leased lines
- Frame Relay
- Asynchronous Transfer Mode (ATM)
- ATM and Frame Relay Service Inter-Working (SIW)
- Multiprotocol Label Switching (MPLS) Virtual Private Network (VPN)
- Voice and Video Enabled IP Security Protocol (IPSec) VPN (V3PN)

Figure 1-3 A Distributed Call Processing Deployment



Best Practices for the Multi-Site Model with Distributed Call Processing

A multi-site WAN with distributed call processing has many of the same requirements as a single site and a multi-site WAN with centralized call processing. Follow the best practices from these other models in addition to the ones listed here for the distributed call processing model. (See Single Site, page 1-2, and Multi-Site WAN with Centralized Call Processing, page 1-4.)

A gatekeeper is one of the key elements in the multi-site WAN model with distributed call processing. A gatekeeper is an H.323 device that provides call admission control and E.164 dial plan resolution. The following best practices apply to the use of a gatekeeper:

- Use a logical hub-and-spoke topology for the gatekeeper. A gatekeeper can manage the bandwidth into and out of a site, or between zones within a site, but it is not aware of the topology.
- 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.
- Use only one type of codec on the WAN because the H.323 specification does not allow for Layer 2, IP, User Data Protocol (UDP), or Real-time Transport Protocol (RTP) header overhead in the bandwidth request. (Header overhead is allowed only in the payload or encoded voice part of the packet.) Using one type of codec on the WAN simplifies capacity planning by eliminating the need to over-provision the IP WAN to allow for the worst-case scenario.
- Gatekeeper networks can scale to hundreds of sites, and the design is limited only by the hub-and-spoke topology.

For more information on gatekeepers, see Gatekeeper Considerations, page 6-10.

Call Admission Control for Distributed Call Processing

Multi-site deployments require some form of call admission control to ensure the voice quality of calls transmitted across network links that have limited available bandwidth. For a multi-site WAN deployment with distributed call processing, use one of the following methods of call admission control, depending on your requirements:

- For Cisco CallManager clusters deployed across a high-speed LAN or MAN, use an Intercluster Trunk, page 1-12.
- For toll bypass or for integration with an existing H.323 environment, use an H.225 Gatekeeper-Controlled Trunk, page 1-13.
- For a WAN with a hub-and-spoke (star) topology, use an Intercluster Gatekeeper-Controlled Trunk, page 1-14.
- For all other cases, including networks that do not follow a hub-and-spoke topology, use an Intercluster Gatekeeper-Controlled Trunk with Locations, page 1-15.

Intercluster Trunk

Cisco CallManager supports three types of trunks for intercluster and H.323 communications:

- Intercluster trunks
- H.225 Gatekeeper-Controlled Trunk, page 1-13
- Intercluster Gatekeeper-Controlled Trunk, page 1-14

The intercluster trunk enables Cisco CallManager clusters to communicate with each other. To use it, configure the intercluster trunk on each Cisco CallManager cluster connected to it.

Note Call admission control is *not* available between clusters using an intercluster trunk; therefore, you should use intercluster trunks only on high-speed LANs or MANs with plenty of available bandwidth.

H.225 Gatekeeper-Controlled Trunk

The H.225 gatekeeper-controlled trunk allows Cisco CallManager to communicate with other Cisco CallManager clusters and with H.323 devices registered to an H.323 gatekeeper. The H.225 gatekeeper-controlled trunk is not recommended in a pure Cisco CallManager environment.

Follow these guidelines when using an H.225 gatekeeper-controlled trunk for call admission control:

- Configure the gatekeeper the same way in each Cisco CallManager cluster.
- Configure the H.225 gatekeeper-controlled trunk in the Cisco CallManager cluster.
- Each Cisco CallManager in a cluster registers an H.225 gatekeeper-controlled trunk with the gatekeeper.
- Calls are load-balanced across the registered trunks in the Cisco CallManager cluster.
- Cisco CallManager supports multiple gatekeepers and trunks.
- Configure a separate zone in the gatekeeper for each site supporting Cisco CallManagers or voice gateways.
- Use the **bandwidth interzone** command on the gatekeeper to control bandwidth between Cisco CallManager clusters and H.323 devices registered directly with the gatekeeper. (See Table 1-1 for bandwidth settings by codec type.)
- A single Cisco IOS gatekeeper can support up to 100 Cisco CallManager clusters.

Example 1-1 illustrates a typical gatekeeper configuration.

Example 1-1 Typical Gatekeeper Configuration for H.225 Gatekeeper-Controlled Trunk

```
gatekeeper
 zone local GK-Headquarters customer.com 10.1.10.100
 zone local GK-BranchA  customer.com
 zone local GK-BranchB customer.com
 zone prefix GK-Headquarters 408…….
 zone prefix GK-BranchA 212…….
 zone prefix GK-BranchB 818…….
 bandwidth interzone GK-Headquarters 200
 bandwidth interzone GK-BranchA 200
 bandwidth interzone GK-BranchB 200
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

Example 1-1 illustrates the following points:

- The **zone local** commands create the gatekeeper zones. Each Cisco CallManager and gateway registers with its configured zone.
- The **zone prefix** is used to route calls between zones.
- The **bandwidth interzone** command allocates the amount of bandwidth available between zones.
- The **gw-type-prefix 1# default technology** command routes unresolved calls within a zone to the device with a registered technology prefix of 1#, which in this example configuration is the Cisco CallManager trunk.
- The **arq reject-unknown-prefix** command prevents call routing loops on redundant Cisco CallManager trunks.

Intercluster Gatekeeper-Controlled Trunk

The intercluster gatekeeper-controlled trunk enables Cisco CallManager to communicate with other Cisco CallManager clusters registered to an H.323 gatekeeper. Cisco recommends that you use the intercluster gatekeeper-controlled trunk in distributed call processing environments.

Follow these guidelines when using an intercluster gatekeeper-controlled trunk:

- Configure the gatekeeper the same way in each Cisco CallManager cluster.
- Configure the intercluster gatekeeper-controlled trunk the same way in each Cisco CallManager cluster.
- Each Cisco CallManager in a cluster registers the intercluster gatekeeper-controlled trunk with the gatekeeper.
- Calls are load-balanced across the registered trunks in the Cisco CallManager cluster.
- Cisco CallManager supports multiple gatekeepers and trunks.
- Configure a separate zone in the gatekeeper for each Cisco CallManager cluster.
- Use the **bandwidth interzone** command on the gatekeeper to control bandwidth between Cisco CallManager clusters and H.323 devices registered directly with the gatekeeper. (See Table 1-1 for bandwidth settings by codec type.)
- A single Cisco IOS gatekeeper can support up to 100 Cisco CallManager clusters.
- You can provide gatekeeper redundancy by using gatekeeper clustering (alternate gatekeeper) or Cisco Hot Standby Router Protocol (HSRP). Use HSRP only if gatekeeper clustering is not available in your software feature set.

Example 1-2 illustrates a typical gatekeeper configuration.

Example 1-2 Typical Gatekeeper Configuration for Intercluster Gatekeeper-Controlled Trunk

```
gatekeeper
 zone local GK-Headquarters customer.com 10.1.10.100
 zone local GK-BranchA  customer.com
 zone local GK-BranchB customer.com
 zone prefix GK-Headquarters 408…….
 zone prefix GK-BranchA 212…….
 zone prefix GK-BranchB 818…….
 bandwidth interzone GK-Headquarters 200
 bandwidth interzone GK-BranchA 200
 bandwidth interzone GK-BranchB 200
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

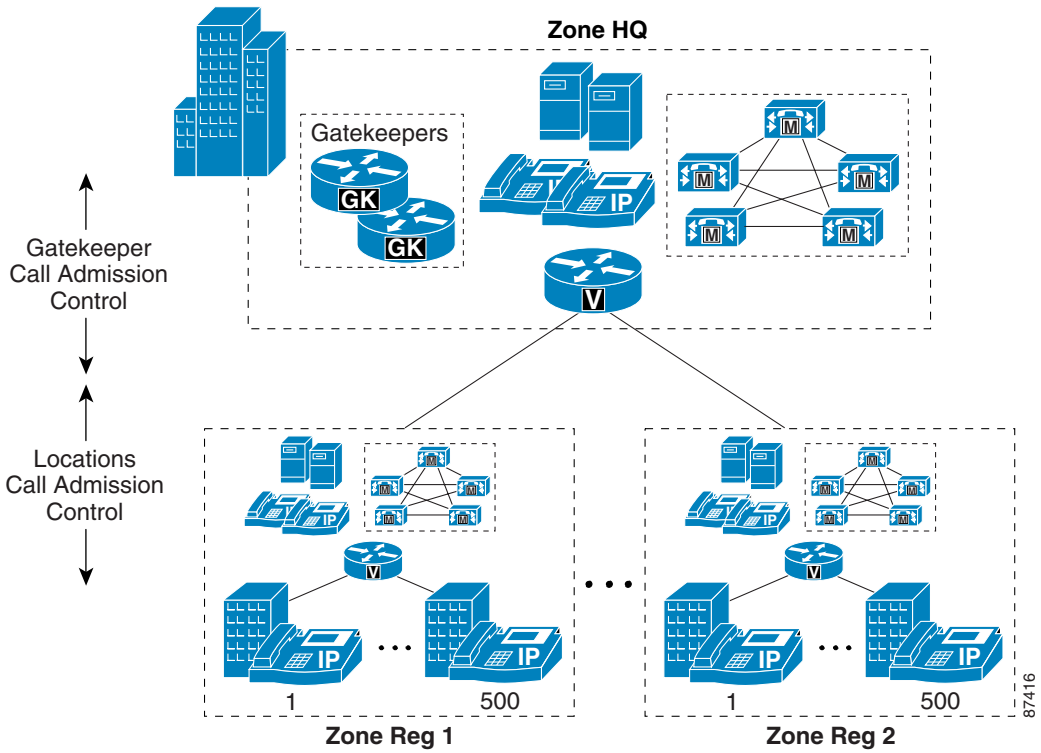
Example 1-2 illustrates the following points:

- The **zone local** commands create the gatekeeper zones. Each Cisco CallManager registers an intercluster gatekeeper-controlled trunk with its configured zone.
- The **zone prefix** is used to route calls between zones.
- The **bandwidth interzone** command allocates the amount of bandwidth available between zones.
- The **gw-type-prefix 1# default technology** command routes unresolved calls within a zone to the device with a registered technology prefix of 1#, which in this example configuration is the Cisco CallManager trunk.
- The **arq reject-unknown-prefix** command prevents call routing loops on redundant Cisco CallManager trunks.

Intercluster Gatekeeper-Controlled Trunk with Locations

In general, the call admission control mechanisms are limited to a hub-and-spoke topology in both centralized and distributed call processing environments. To provide call admission control in deployments that do not use a hub-and-spoke topology, you can combine the locations and gatekeeper mechanisms as illustrated in Figure 1-4.

Figure 1-4 Combining Locations and Gatekeeper Call Admission Control



Follow these guidelines when combining an intercluster gatekeeper-controlled trunk with locations for call admission control:

- Use locations-based call admission control for sites with no local Cisco CallManager.
- Use gatekeeper-based call admission control between Cisco CallManager clusters.
- For each site without a local Cisco CallManager, configure a location for that site in the Cisco CallManager cluster supporting the site.
- Configure the appropriate bandwidth limit for each site according to the type of codec used at that site. (See Table 1-1 for bandwidth settings.)
- Assign each device configured in Cisco CallManager to a location. If you move a device to another location, change its location configuration as well.
- Cisco CallManager supports up to 500 locations.
- Each Cisco CallManager registers a intercluster gatekeeper-controlled trunk with the gatekeeper.
- Configure the gatekeeper the same way in each Cisco CallManager cluster.
- Configure the intercluster gatekeeper-controlled trunk the same way in each Cisco CallManager cluster.

- Calls are load-balanced across the registered trunks in the Cisco CallManager cluster.
- Cisco CallManager supports multiple gatekeepers and trunks.
- Configure a separate zone in the gatekeeper for each Cisco CallManager cluster.
- Use the **bandwidth interzone** command on the gatekeeper to control bandwidth between Cisco CallManager clusters and H.323 devices registered directly with the gatekeeper. (See Table 1-1 for bandwidth settings by codec type.)
- A single Cisco IOS gatekeeper can support up to 100 Cisco CallManager clusters.
- You can provide gatekeeper redundancy by using gatekeeper clustering (alternate gatekeeper) or Cisco Hot Standby Router Protocol (HSRP). Use HSRP only if gatekeeper clustering is not available in your software feature set.

Example 1-3 illustrates a typical gatekeeper configuration.

Example 1-3 Typical Gatekeeper Configuration for Intercluster Gatekeeper-Controlled Trunk with Locations

```
gatekeeper
 zone local GK-HQ customer.com 10.1.10.100
 zone local GK-Reg1  customer.com
 zone local GK-Reg2 customer.com
 zone prefix GK-HQ 408*
 zone prefix GK-Reg1 718*
 zone prefix GK-Reg1 212*
 zone prefix GK-Reg2 818*
 zone prefix GK-Reg2 602
 bandwidth interzone GK-HQ 768
 bandwidth interzone GK-Reg1 768
 bandwidth interzone GK-Reg2 768
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

Example 1-3 illustrates the following points:

- The **zone local** commands create the gatekeeper zones. Each Cisco CallManager registers an intercluster gatekeeper-controlled trunk with its configured zone.
- The **zone prefix** is used to route calls between zones. You may define multiple zone prefixes for the same zone, if needed.
- The **bandwidth interzone** command allocates the amount of bandwidth available between zones.
- The **gw-type-prefix 1# default technology** command routes unresolved calls within a zone to the device with a registered technology prefix of 1#, which in this example configuration is the Cisco CallManager trunk.
- The **arq reject-unknown-prefix** command prevents call routing loops on redundant Cisco CallManager trunks.

Clustering Over the IP WAN

You may deploy a single Cisco CallManager cluster across multiple sites that are connected by an IP WAN with QoS features enabled. This section provides a brief overview of clustering over the WAN. For further information, refer to the section on Call Processing, page 6-1.

Clustering over the WAN can support two types of deployments:

- Local Failover Deployment Model, page 1-17

  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 or three sites with Cisco CallManager servers and a maximum of 5000 or 2500 IP phones per site, respectively. This model allows for up to 10,000 IP phones in the two-site configuration and 7,500 IP phones in the three-site configuration.

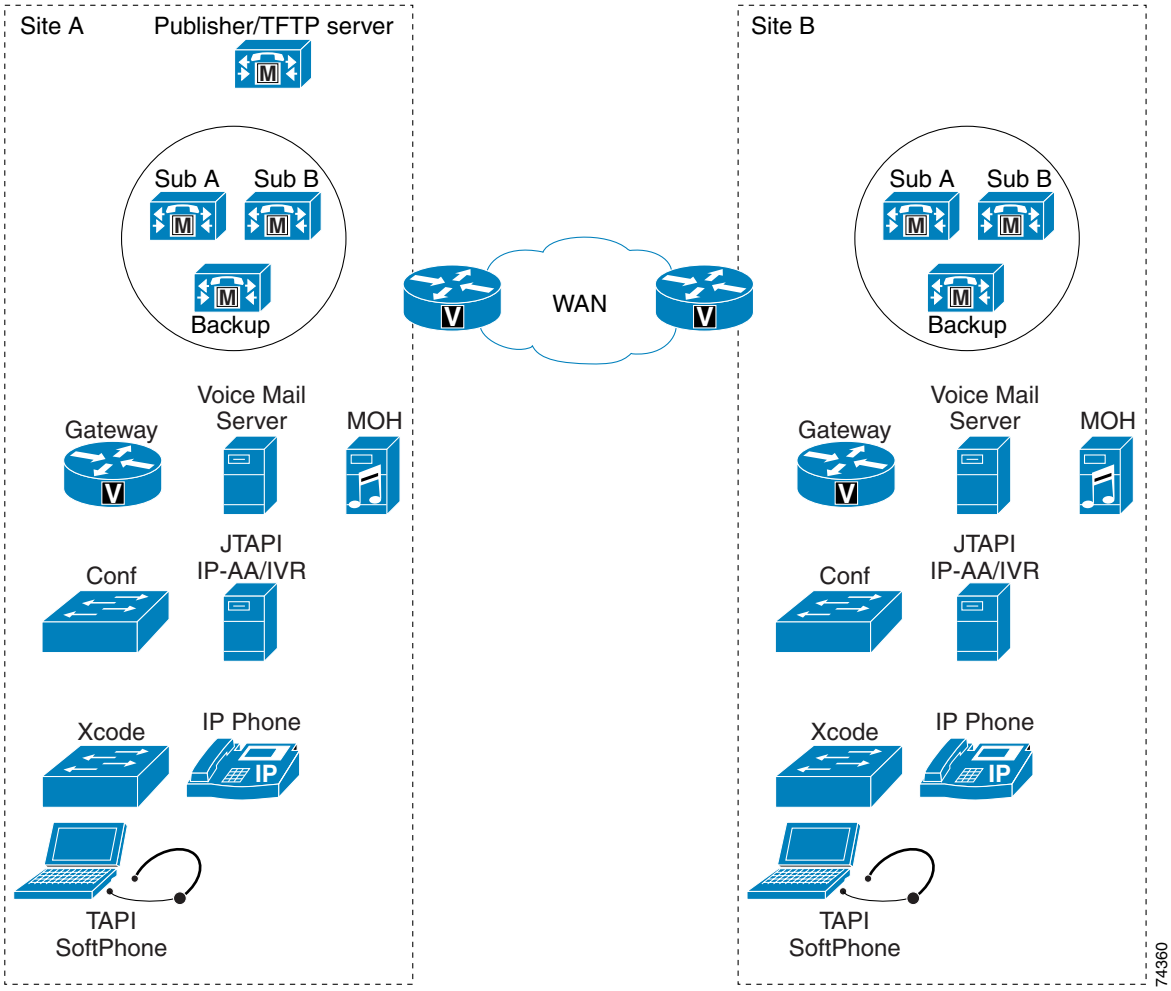
- Remote Failover Deployment Model, page 1-19

  Remote failover allows you to deploy the backup servers over the WAN. Using this deployment model, you may have up to six sites with Cisco CallManager subscribers and one or two sites containing the Cisco CallManager backup server. This deployment allows for up to 10,000 IP phones shared over the required number of sites.

Local Failover Deployment Model

The local failover deployment model provides the most resilience for clustering over the WAN. Each of the sites in this model contains at least one primary Cisco CallManager subscriber and one backup subscriber. This configuration allows for either a two-site deployment with 5000 IP phones per site or a three-site deployment with 2500 IP phones per site. (See Figure 1-5.)

Figure 1-5 Local Failover Model with Two Sites



In summary, observe the following guidelines when implementing the local failover model:

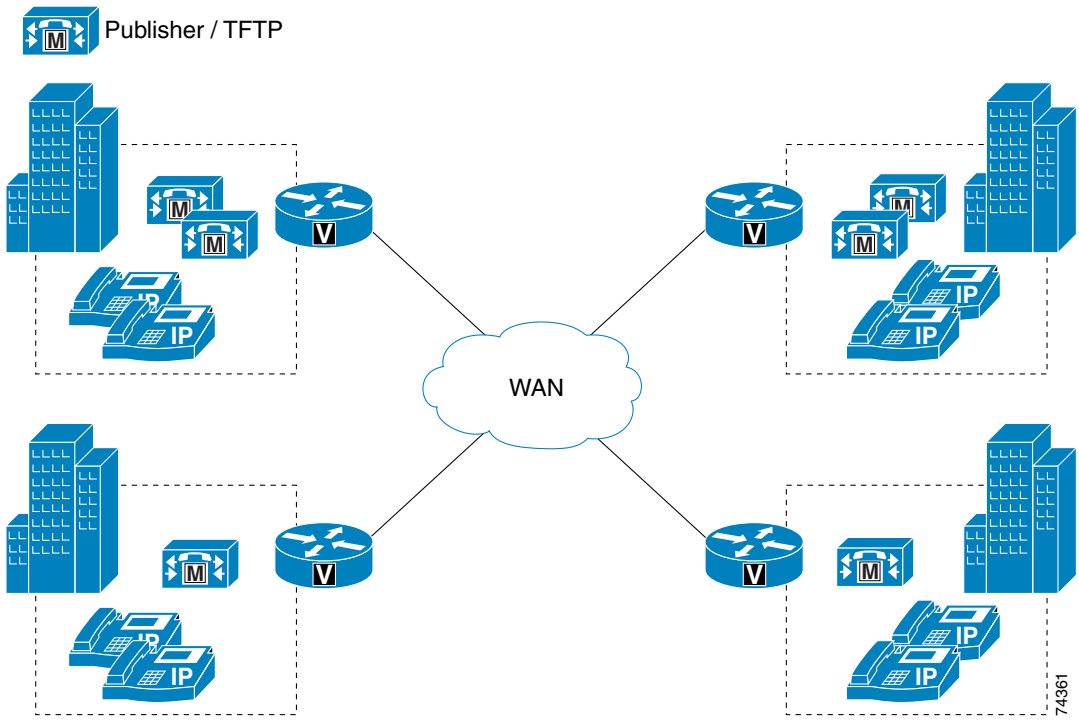
- Configure each site to contain at least one primary Cisco CallManager subscriber and one backup subscriber.
- Configure Cisco CallManager *groups* and *device pools* to allow devices within the site to register with only the servers at that site under all conditions.
- Cisco highly recommends that 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 could also extend this practice to include a voice mail system at each site. Under a WAN failure condition, only sites without access to the publisher database might loose a small amount of functionality.
- Every 10,000 busy hour call attempts (BHCA) in the cluster requires 900 kbps of bandwidth for Intra-Cluster Communication Signaling (ICCS). This is a minimum bandwidth requirement, and bandwidth is allocated in multiples of 900 kbps.

- In addition to the real-time ICCS bandwidth, intra-cluster bandwidth is required for the SQL, CTI Manager, and LDAP traffic. The amount of additional bandwidth is dependant on the use of the system. For instance, the use of Extension Mobility increases the amount of SQL traffic between servers.
- A maximum Round Trip Time (RTT) of 40 ms is allowed between any two servers in the Cisco CallManager cluster. This time equates to a 20 ms maximum one-way delay, or a transmission distance of approximately 1860 miles (3000 km) under ideal conditions.
- The local failover model requires Cisco CallManager Release 3.1 or later.

Remote Failover Deployment Model

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 three to six sites with IP phones and other devices normally registered with a maximum of four servers. (See Figure 1-6.)

Figure 1-6 Remote Failover Model with Four Sites



In summary, observe the following guidelines when implementing the remote failover model:

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

- 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 with IP phones to provide the highest level of resiliency. You could also extend this practice to include a voice mail system at each site. Under a WAN failure condition, only sites without access to the publisher database might loose a small amount of functionality.
- Every 10,000 busy hour call attempts (BHCA) in the cluster requires 900 kbps of bandwidth for Intra-Cluster Communication Signaling (ICCS). This is a minimum bandwidth requirement, and bandwidth is allocated in multiples of 900 kbps.
- In addition to the real-time ICCS bandwidth, intra-cluster bandwidth is required for the SQL, CTI Manager, and LDAP traffic. The amount of additional bandwidth is dependant on the use of the system. For instance, the use of Extension Mobility increases the amount of SQL traffic between servers.
- Signaling or Control Plane traffic requires additional bandwidth when devices are registered across the WAN with a remote Cisco CallManager server in the same cluster.
- A maximum Round Trip Time (RTT) of 40 ms is allowed between any two servers in the Cisco CallManager cluster. This time equates to a 20 ms maximum one-way delay, or a transmission distance of approximately 1860 miles (3000 km) under ideal conditions.
- The remote failover model requires Cisco CallManager Release 3.1 or later.

Call Admission Control for Clustering Over the IP WAN

If calls are allowed across the WAN between sites, then you must provide call admission control between those sites by configuring Cisco CallManager *locations* for those sites in addition to the default location for the other sites. Even if the bandwidth is over-provisioned for the number of devices, it is still best to configure call admission control based on locations.

Multi-Site MPLS WAN Considerations

This section explains how to adapt the call admission control mechanisms, described previously in this chapter, to multi-site deployments with Multiprotocol Label Switching (MPLS) WANs.

The main design difference between traditional Layer 2 WAN technologies and MPLS is that an MPLS WAN does not conform to a hub-and-spoke topology but instead provides full-mesh connectivity between all sites. This topology difference has implications on the call admission control mechanisms that must be adopted.

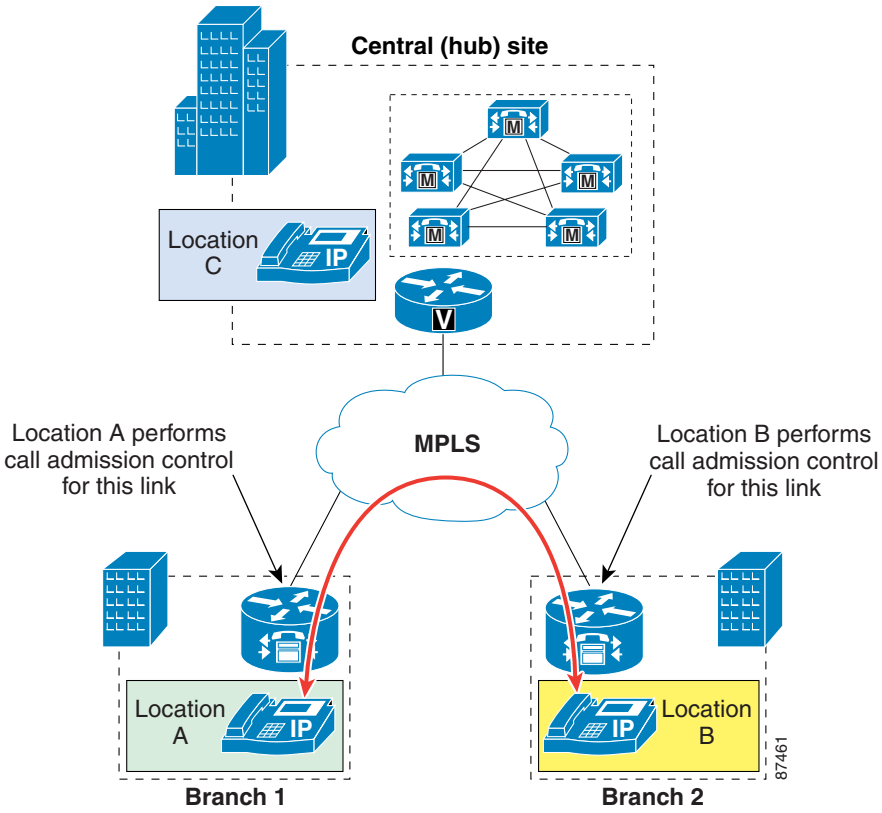
Purely Centralized Deployments

In single-cluster centralized call processing deployments, the call admission control function is performed by the *locations* construct within Cisco CallManager.

In a hub-and-spoke WAN topology (for example, Frame Relay or ATM), each link to and from a branch site terminates at the central site. For example, in a Frame Relay network, all Permanent Virtual Circuits (PVCs) from the branch routers are aggregated at the central site's head-end router. In such a scenario, there is no need to apply call admission control to devices at the central site because the bandwidth accounting occurs at the branch ends of the WAN links. Therefore, within the Cisco CallManager Locations configuration, devices at the central site are left in the <None> location, while devices at each branch are placed in their appropriate location to ensure proper call admission control.

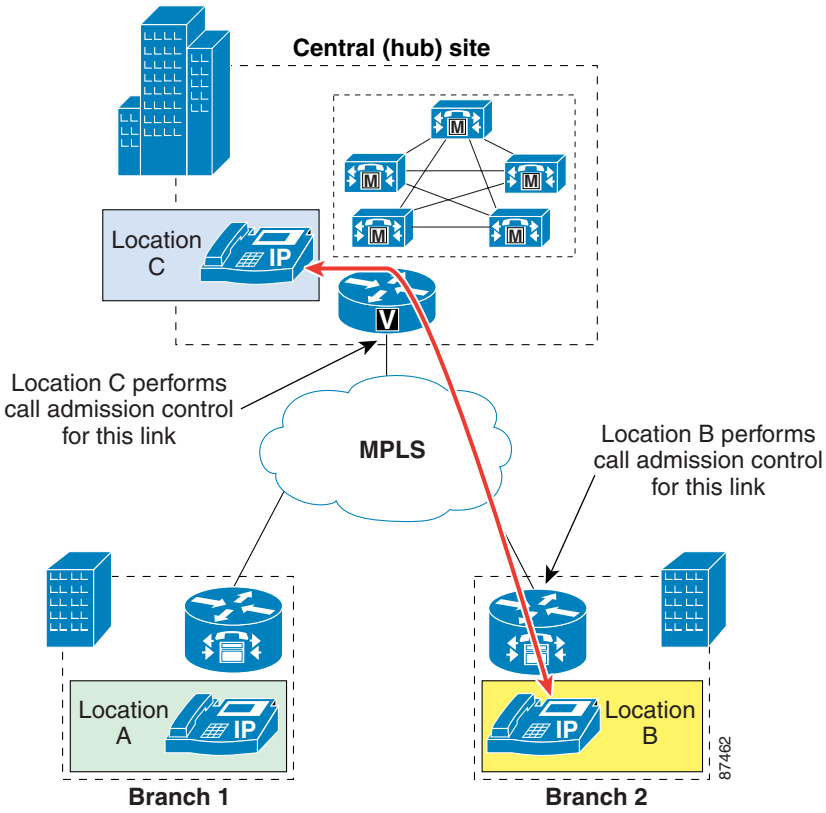
With an MPLS WAN network, all branches are deemed to be adjacent at Layer 3, thus they do not have to rely on the central site for connectivity. Figure 1-7 illustrates a spoke-to-spoke call between two branch sites in this type of deployment.

Figure 1-7 Spoke-to-Spoke Calls in an MPLS Deployment



Also, in an MPLS WAN, the link connecting the central site to the WAN does not aggregate every branch's WAN link. By placing 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 branch links. (See Figure 1-8.)

Figure 1-8 Calls to and from the Hub in an MPLS Deployment

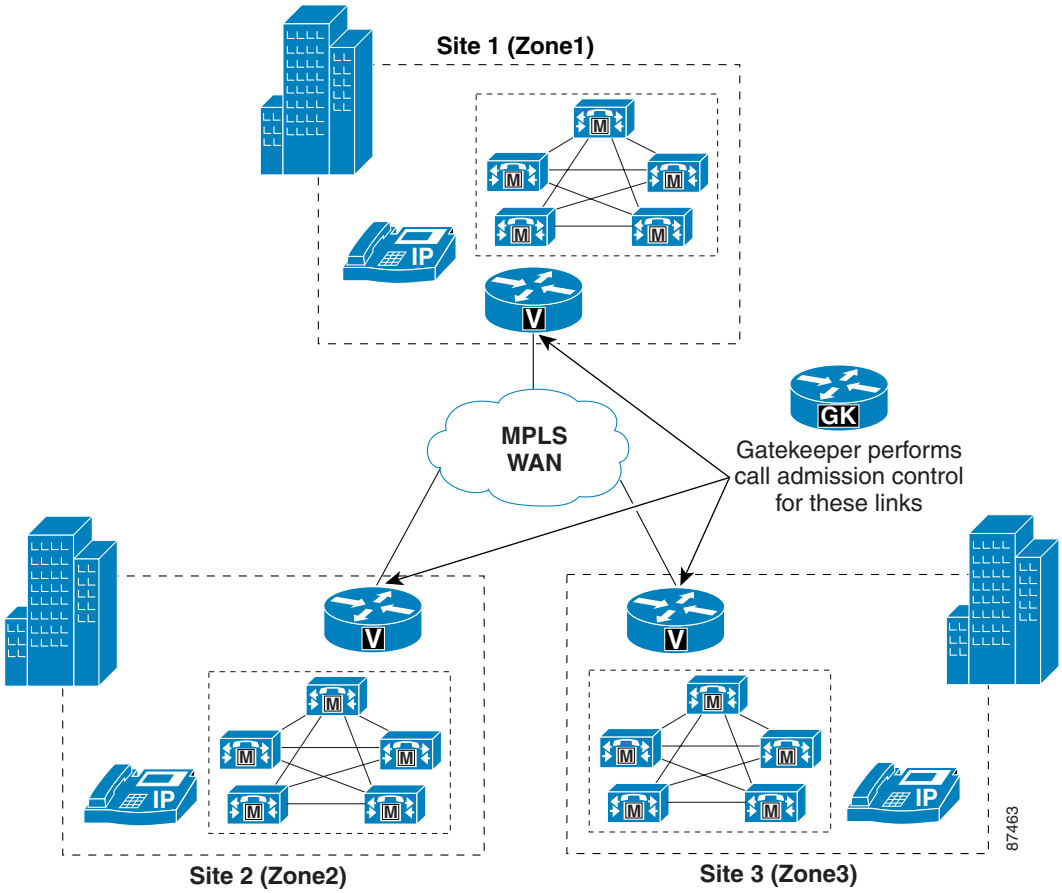


When all the available bandwidth for a particular site has been utilized, you can provide automatic failover to the PSTN using the automated alternate routing (AAR) feature within Cisco CallManager. (For more information on AAR, see the section on Automated Alternate Routing, page 7-9.)

Purely Distributed Deployments

In multi-site deployments where a Cisco CallManager cluster is present at each site and the sites are linked through an MPLS WAN, a gatekeeper can provide call admission control between the sites, with each site being placed in a different gatekeeper zone. This is the same mechanism adopted for hub-and-spoke topologies based on Layer 2 WAN technologies. (See Figure 1-9.)

Figure 1-9 Gatekeeper Call Admission Control in a Distributed Deployment with MPLS



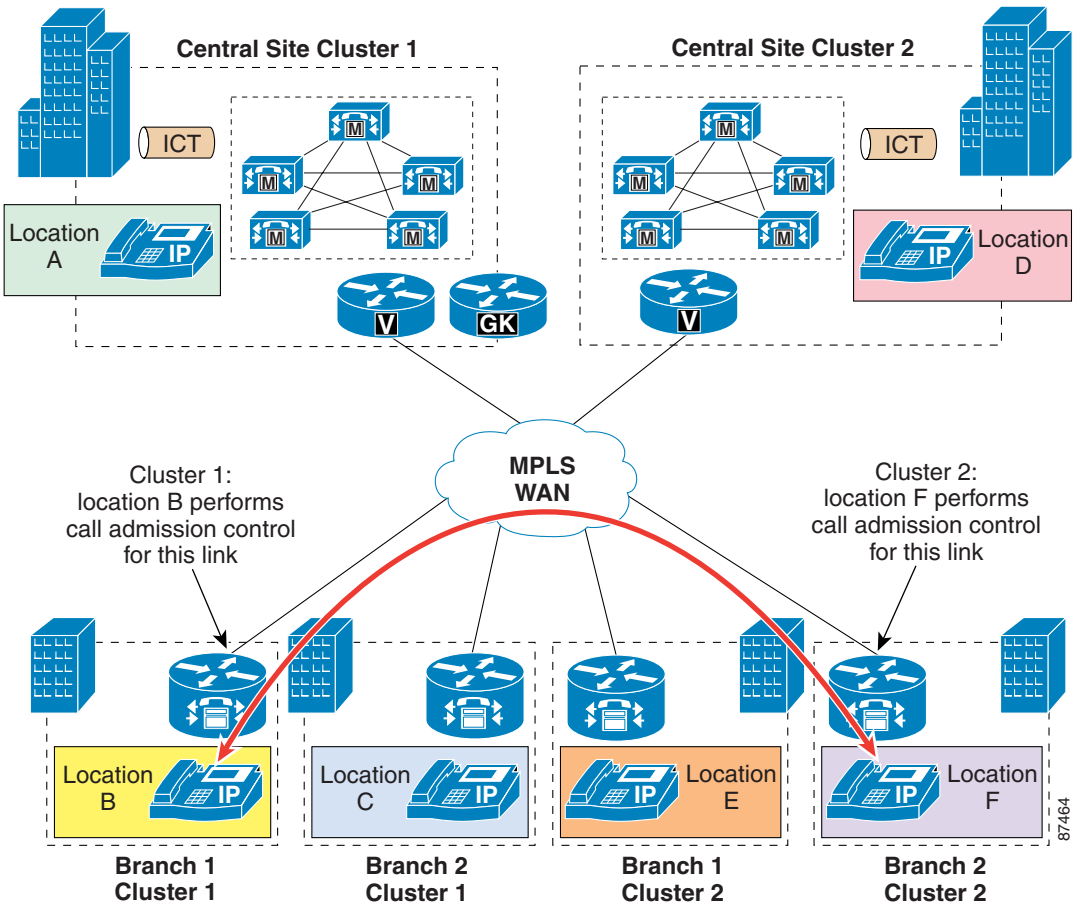
When all the available bandwidth for a particular site has been utilized, you can provide automatic failover to the PSTN using the route list and route group construct for the route patterns that connect each cluster to the gatekeeper.

Hybrid Centralized/Distributed Deployments

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

When calls occur between branches belonging to different clusters, the audio path is established between the two branches directly, with no media transiting through each cluster's central site. Therefore, call admission control is required only on the WAN links at the two branches. (See Figure 1-10.)

Figure 1-10 Multiple Clusters Connected by Intercluster Trunks (ICT)



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 intercluster trunks are purely signaling devices, and there is no media transiting through them. Therefore, all intercluster 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.

When all the available bandwidth 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 cluster (For more information on AAR, see the section on Automated Alternate Routing, page 7-9.)

Multi-Cluster Campus TFTP Services

Cisco IP Telephony devices rely on Trivial File Transfer Protocol (TFTP) to download the files required for their proper initialization, configuration, and registration. The TFTP server is part of the Cisco CallManager cluster and is responsible for creating and transferring configuration files. In a single-cluster environment, all devices rely on a single TFTP server. The TFTP server address is typically configured as option 150 in the DHCP server's scope.

In multi-cluster systems, it is possible to have a single subnet or VLAN containing phones from multiple clusters. In this situation, the TFTP server whose address is provided to all phones in the subnet or VLAN must service the file transfer requests made by each phone, regardless of which cluster contains the phone. This centralized TFTP server, therefore, must have access to files created and managed by other clusters.

To provide access to files in other clusters, each cluster's TFTP server must be configured to create and manage configuration files on the centralized TFTP server's drive. To configure this file access, use the alternate file location entry under each TFTP server's configuration (with the exception of the centralized TFTP server's cluster).

Note With Cisco CallManager Release 3.2 and later, Cisco TFTP servers cache the configuration files for the IP phones in RAM by default. For those files to be written to a centralized TFTP server, you must disable caching by modifying the following service parameters from their default values on each TFTP server configured to write to the centralized TFTP server:

- Enable Caching of Configuration Files: **False** (required)
- Enable Caching of Constant and Bin Files at Startup: **False** (recommended)

If the TFTP server receives a request for a file that it does not have (such as a configuration file created and maintained by the TFTP server in a different cluster), it will search for the file in a list of alternate file locations. You have to configure the centralized TFTP server to search through the subdirectories associated with the other clusters.

Example 1-4 Alternate TFTP File Locations

A large campus system is deployed using four clusters, and each cluster contains a TFTP server. In all subnets, the DHCP scope provides the IP address of TFTP1 as option 150. Assume that TFTP1 is the centralized TFTP server for this system of four clusters. The other servers, TFTP2, TFTP3 and TFTP4, are each configured to write their configuration files to the following subdirectories on TFTP1's drive:

- TFTP2's alternate file location is set to: //TFTP1/Program Files/Cisco/TFTPpath/TFTP2.
- TFTP3's alternate file location is set to: //TFTP1/Program Files/Cisco/TFTPpath/TFTP3.
- TFTP2's alternate file location is set to: //TFTP1/Program Files/Cisco/TFTPpath/TFTP4.

TFTP1 is configured to search in alternate locations as follows:

- Alternate File Location 1: c:\Program Files\Cisco\TFTPpath\TFTP2
- Alternate File Location 2: c:\Program Files\Cisco\TFTPpath\TFTP3
- Alternate File Location 3: c:\Program Files\Cisco\TFTPpath\TFTP4

Note In this example, TFTP1 represents the IP address of TFTP1. Also, TFTP1 requires that NT subdirectories be created manually for TFTP2, TFTP3 and TFTP4.

Cisco recommends that you use Universal Naming Convention (UNC) paths (in the format \\<*IP_address*>\<*Full_path_to_folder*>) to point a TFTP server to alternate file locations. Cisco does *not* recommend creating non-default NT "shares" or using DNS names. Also, ensure that all clusters meet the proper login ID, password, and security privileges (workgroup, domain, or directory-based) for the Cisco TFTP service.

Also, in large campus deployments, adjust the Maximum Serving Count service parameter as follows: The suggested value for a dedicated TFTP server is 1500 for a single processor system and 3000 for a dual processor system. If the dual processor system is running Windows 2000 Advanced Server, the serving count can be up to 5000.

Redundancy

Option 150 allows more than a single IP address to be returned to phones as part of the DHCP scope. The phone tries the first address in the list, and it tries a subsequent address only if it cannot establish communications with the first TFTP server. This address list provides a redundancy mechanism that enables phones to obtain TFTP services from another server even if their primary TFTP server has failed.

Two TFTP servers can be configured in a cluster, and each can create and manage separate lists of the same configuration files. In a multi-cluster environment, two separate groups of TFTP servers can be configured to ensure redundancy, each with a member serving as the centralized server.

If we wanted to provide TFTP redundancy for the case described in Example 1-4, we could configure each cluster with two TFTP servers. All primary TFTP servers would be configured to write their configuration files to TFTP1_P, while all the secondary TFTP servers would write theirs to TFTP1_S, as follows:

- TFTP2_P's alternate file location is set to: //TFTP1_P/Program Files/Cisco/TFTPpath/TFTP2.
- TFTP3_P's alternate file location is set to: //TFTP1_P/Program Files/Cisco/TFTPpath/TFTP3.
- TFTP2_P's alternate file location is set to: //TFTP1_P/Program Files/Cisco/TFTPpath/TFTP4.
- TFTP2_S's alternate file location is set to: //TFTP1_S/Program Files/Cisco/TFTPpath/TFTP2.
- TFTP3_S's alternate file location is set to: //TFTP1_S/Program Files/Cisco/TFTPpath/TFTP3.
- TFTP2_S's alternate file location is set to: //TFTP1_S/Program Files/Cisco/TFTPpath/TFTP4.

Both TFTP1_P and TFTP1_S need to be configured as in Example 1-4 to search through the list of alternate file locations.

Load Balancing

The preceding sections explain how to use one TFTP server at a time to service phones from multiple clusters. For this approach, Cisco recommends that you grant different ordered lists of TFTP servers to different subnets, to allow for load balancing. For example:

- In subnet 10.1.1.0/24, Option 150 is set to: TFTP1, TFTP2
- In subnet 10.1.2.0/24, Option 150 is set to: TFTP2, TFTP1

Under normal operation, a phone in subnet 10.1.1.0/24 will request TFTP services from TFTP1, while a phone in subnet 10.1.2.0/24 will request TFTP services from TFTP2. If TFTP1 fails, then phones from both subnets will request TFTP services from TFTP2.

Load balancing avoids having a single TFTP server "hot spot", where all phones from multiple clusters rely on the same server for service. TFTP load balancing is especially important when phone software loads are transferred, such as during a Cisco CallManager upgrade, because more files of larger size are being transferred, thus imposing a bigger load on the TFTP server.

Design Considerations for Section 508 Conformance

Regardless of which deployment model you choose, you should consider designing your IP telephony network to make the telephony features more accessible to users with disabilities, in conformance with Section 255 of the Telecommunications Act and U.S. Section 508.

Observe the following basic design guidelines when configuring your IP telephony network to conform to Section 508:

- Enable Quality of Service (QoS) on the network.
- Configure only the G.711 codec for phones that will be connected to a terminal teletype (TTY) device or a Telephone Device for the Deaf (TDD). Although low bit-rate codecs such as G.729 are acceptable for audio transmissions, they do not work well for TTY/TDD devices if they have an error rate higher than 1% Total Character Error Rate (TCER).
- Configure TTY/TDD devices for G.711 across the WAN, if necessary.
- Enable (turn ON) Echo Cancellation for optimal performance.
- Voice Activity Detection (VAD) does not appear to have an effect on the quality of the TTY/TDD connection, so it may be disabled or enabled.
- Configure the appropriate *regions* and *device pools* in Cisco CallManager to ensure that the TTY/TDD devices always use G.711 codes.
- Connect the TTY/TDD to the IP telephony network in either of the following ways:
  - Direct connection (Recommended method)

    Plug a TTY/TDD with an RJ-11 analog line option directly into a Cisco FXS port. Any FXS port will work, such as the one on a Cisco IP Phone 7900 Series or on the Cisco VG 248, Catalyst 6000, Cisco ATA 186/188 modules, or any other Cisco voice gateway with an FXS port. Cisco recommends this method of connection.
  - Acoustic coupling

    Place the IP phone handset into a coupling device on the TTY/TDD. Acoustic coupling is less reliable than an RJ-11 connection because the coupling device is generally more susceptible to transmission errors caused by ambient room noise and other factors.
- If stutter dial tone is required, use an analog phone in conjunction with an FXS port on the Cisco VG 248 or ATA 186/188.

Network Infrastructure

This chapter describes requirements of the network infrastructure needed to build an IP telephony solution in an enterprise campus environment. Figure 2-1 illustrates the roles of the various devices that form the network infrastructure, and Table 2-1 summarizes the features required for each of these roles.

Figure 2-1 Typical Campus Network Infrastructure

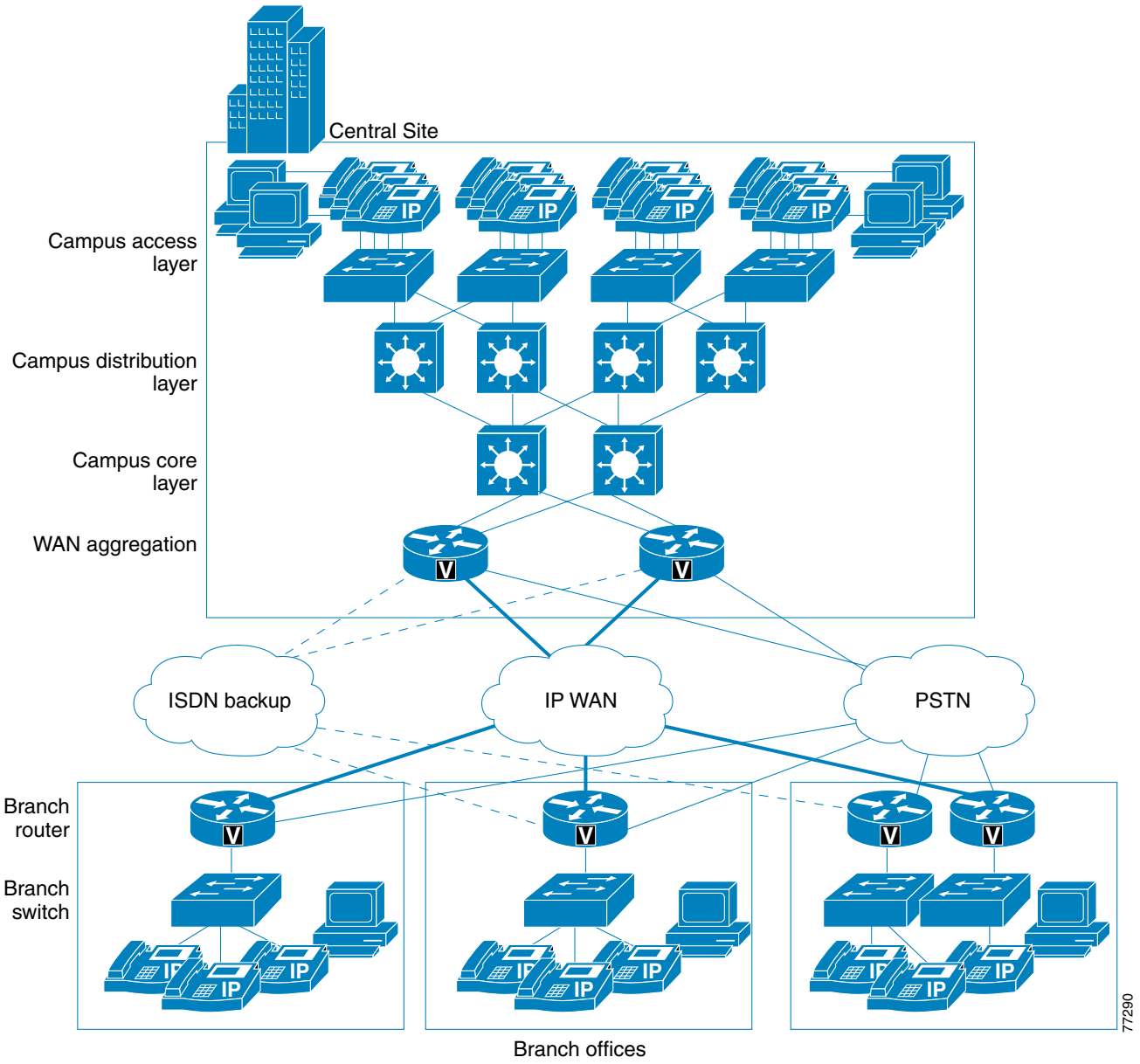


Table 2-1 Required Features for Each Role in the Network Infrastructure

Infrastructure Role	Required Features
Campus Access Switch	• In-Line Power • Multiple Queue Support • 802.1p and 802.1Q • Fast Link Convergence
Campus Distribution or Core Switch	• Multiple Queue Support • 802.1p and 802.1Q • Traffic Classification • Traffic Reclassification
WAN Aggregation Router (Site that is at the hub of the network)	• Multiple Queue Support • Traffic Shaping • Link Fragmentation and Interleaving (LFI) • Link Efficiency • Traffic Classification • Traffic Reclassification • 802.1p and 802.1Q
Branch Router (Spoke site)	• Multiple Queue Support • LFI • Link Efficiency • Traffic Classification • Traffic Reclassification • 802.1p and 802.1Q
Branch or Smaller Site Switch	• In-Line Power • Multiple Queue Support • 802.1p and 802.1Q

LAN Infrastructure

Table 2-2 lists the traffic classification requirements for the LAN infrastructure.

Table 2-2 Traffic Classification Guidelines for Various Types of AVVID Network Traffic

Traffic Type	Layer 2 Class of Service (CoS)	Layer 3 IP Precedence	Layer 3 Differentiated Services Code Point (DSCP)
Voice Real-Time Transport Protocol (RTP)	5	5	EF
Voice control signaling	3	3	AF31
Video	4	4	AF41
Data	0, 1, 2	0, 1, 2	0 to AF23

WAN Infrastructure

Table 2-3 lists the QoS features and tools required for the WAN infrastructure.

Table 2-3 QoS Features and Tools Required to Support IP Telephony for each WAN Technology and Link Speed

WAN Technology	Link Speed: 56 kbps to 768 kbps	Link Speed: Greater than 768 kbps
Leased Lines	• Multilink Point-to-Point Protocol (MLP) • MLP Link Fragmentation and Interleaving (LFI) • Low Latency Queuing (LLQ) • Optional: Compressed Real-Time Transport Protocol (cRTP)	• LLQ
Frame Relay (FR)	• Traffic Shaping • LFI (FRF.12) • LLQ • Optional: cRTP	• Traffic Shaping • LLQ
Asynchronous Transfer Mode (ATM)	• TX-ring buffer changes • MLP over ATM • MLP LFI • LLQ • Optional: cRTP (requires MLP)	• TX-ring buffer changes • LLQ
Frame Relay and ATM Service Inter-Working (SIW)	• TX-ring buffer changes • MLP over ATM and FR • MLP LFI • LLQ • Optional: cRTP (requires MLP)	• TX-ring buffer changes • MLP over ATM and FR • LLQ

Table 2-3 QoS Features and Tools Required to Support IP Telephony for each WAN Technology and Link Speed (continued)

WAN Technology	Link Speed: 56 kbps to 768 kbps	Link Speed: Greater than 768 kbps
Multiprotocol Label Switching (MPLS)	• Same as above, according to the interface technology • Class-based marking is generally required to remark flows according to service provider specifications	• Same as above, according to the interface technology • Class-based marking is generally required to remark flows according to service provider specifications

Bandwidth Provisioning

You should provision the WAN link bandwidth so that all the major applications combined (voice, video, and data) do not consume more than 75% of the available link bandwidth.

Table 2-4 lists the bandwidth consumed by the voice payload only, at a default packet rate of 50 packets per second (pps).

Table 2-4 Bandwidth Consumption for Voice Payload and IP Header Only

CODEC	Sampling Rate	Voice Payload in Bytes	Packets per Second	Bandwidth per Conversation
G.711	20 ms	160	50.0	80.0 kbps
G.711	30 ms	240	33.3	74.7 kbps
G.729A	20 ms	20	50.0	24.0 kbps
G.729A	30 ms	30	33.3	18.7 kbps

Table 2-5 lists the amount of bandwidth consumed by voice traffic when the Layer 2 headers are included in the calculation.

Table 2-5 Bandwidth Consumption with Layer 2 Headers Included

CODEC	Ethernet 14 Bytes of Header	PPP 6 Bytes of Header	ATM 53-Byte Cells with a 48-Byte Payload	Frame Relay 4 Bytes of Header
G.711 at 50.0 pps	85.6 kbps	82.4 kbps	106.0 kbps	81.6 kbps
G.711 at 33.3 pps	78.4 kbps	76.3 kbps	84.8 kbps	75.7 kbps
G.729A at 50.0 pps	29.6 kbps	26.4 kbps	42.4 kbps	25.6 kbps
G.729A at 33.3 pps	22.4 kbps	20.3 kbps	28.3 kbps	19.7 kbps

The calculations in this section assume an average of 10 calls per hour per phone.

Equation 1: Recommended Bandwidth Needed for Control Traffic, with *No* TAPI Applications.

Bandwidth (bps) = 150 * (Number of IP phones and gateways in the branch)

Equation 2: Recommended Bandwidth Needed for Control Traffic, with TAPI Applications.

Bandwidth with TAPI (bps) = 225 * (Number of IP phones and gateways in the branch)

Table 2-6 summarizes the recommended bandwidths for various branch office sizes.

Table 2-6 Recommended Bandwidth for Call Control Traffic With and Without TAPI Applications

Branch Office Size (Number of IP Phones and Gateways)	Recommended Bandwidth for Control Traffic (no TAPI)	Recommended Bandwidth for Control Traffic (with TAPI)
1 to 30	8 kbps	8 kbps
40	8 kbps	9 kbps
50	8 kbps	11 kbps
60	9 kbps	14 kbps
70	11 kbps	16 kbps
80	12 kbps	18 kbps
90	14 kbps	21 kbps
100	15 kbps	23 kbps
110	17 kbps	25 kbps
120	18 kbps	27 kbps
130	20 kbps	30 kbps
140	21 kbps	32 kbps
150	23 kbps	34 kbps

Note The values in Table 2-6 reflect Layer 3 bandwidth. When provisioning a WAN link, you must add Layer 2 overhead to these numbers according to the Layer 2 technology used.

Advanced Formulas

The previous formulas presented in this section assume an average call rate per phone of 10 calls per hour. However, this rate might not correspond to your deployment if the call patterns are significantly different (for example, with call center agents at the branches). To calculate call control bandwidth requirements in these cases, use the following formulas, which contain an additional variable (CH) that represents the average calls per hour per phone:

Equation 3: Recommended Bandwidth for a Branch with *No* TAPI Applications.

Bandwidth (bps) = (39 + 10.8 ∗ CH) ∗ (Number of IP phones and gateways in the branch)

Equation 4: Recommended Bandwidth for a Remote Branch with TAPI Applications.

Bandwidth with TAPI (bps) = (39 + 18.3 ∗ CH) ∗ (Number of IP phones and gateways in the branch)

Bandwidth Provisioning for Call Control Traffic with Distributed Call Processing

For WAN links that connect distributed Cisco CallManager clusters, the *signaling* component of the required bandwidth can be calculated as follows:

Assuming an average call duration of 2 minutes and 100 percent utilization of each virtual tie line, we can derive that each tie line carries a volume of 30 calls per hour. This assumption allows us to obtain the following formula that expresses the recommended bandwidth for call control traffic as a function of the number of virtual tie lines.

Equation 5: Recommended Bandwidth Based on Number of Virtual Tie Lines.

Recommended Bandwidth (bps) = 116 ∗ (Number of virtual tie lines)

If we take into account the fact that 8 kbps is the smallest bandwidth that can be assigned to a queue on a Cisco IOS router, we can deduce that a minimum queue size of 8 kbps can accommodate the call control traffic generated by *up to 70 virtual tie lines*. This amount should be sufficient for most large enterprise deployments.

Traffic Prioritization

For multi-service traffic over an IP WAN, Cisco recommends low-latency queuing (LLQ). The following prioritization criteria are also recommended:

- The criterion for voice to be placed into a priority queue is the differentiated services code point (DSCP) value of EF, or IP precedence value of 5.
- The criterion for video conferencing traffic to be placed into a priority queue is a DSCP value of AF41, or IP precedence value of 4. Note that one-way video traffic, such as IP/TV, should use a class-based weighted fair queuing scheme because that type of traffic has a much higher delay tolerance.
- Voice control protocols, such as H.323 and Skinny Client Control Protocol (SCCP), require their own class-based weighted fair queue. The entrance criterion for this queue is a DSCP value of AF31, which corresponds to an IP precedence value of 3.

Link Efficiency Techniques

Compressed Real-Time Transport Protocol (cRTP)

You can increase link efficiency further by using Compressed Real-Time Transport Protocol (cRTP). This protocol compresses a 40-byte IP, User Datagram Protocol (UDP), and RTP header into approximately two to four bytes. Use cRTP on a particular link only if that link meets *all* of the following conditions:

- Voice traffic represents more than 30% of the load on the specific link.
- The link uses a low bit-rate codec (such as G.729).
- No other real-time application (such as video conferencing) is using the same link.

For additional recommendations about using cRTP with a Voice and Video Enabled IPSec VPN (V3PN), refer to the V3PN documentation available at

http://cisco.com/go/srnd

Do not use cRTP on ATM and Frame Relay Service Inter-Working (SIW) links.

An important parameter to consider before using cRTP is router CPU utilization, which is adversely affected by compression and decompression operations.

Link Fragmentation and Interleaving (LFI)

For low-speed links (less than 768 kbps), use link fragmentation and interleaving (LFI).

Traffic Shaping

Traffic shaping is required for multiple access, non-broadcast media such as ATM and Frame Relay, where the physical access speed varies between two endpoints, and several branch sites are typically aggregated to a single router interface at the central site.

The following scenarios illustrate why traffic shaping is needed when transporting voice and data traffic on the same IP WAN:

- Line speed mismatch

  While the central site interface is typically a high-speed one (such as T1 or higher), smaller remote branch interfaces may have significantly lower line speeds, such as 64 kbps. If data is sent at full rate from the central site to a slow-speed remote site, the interface at the remote site can become congested and degrade voice performance.

- Remote-to-central-site oversubscription

  It is common practice in Frame Relay or ATM networks to oversubscribe bandwidth when aggregating many remote sites to a single central site. For example, there may be multiple remote sites that connect to the WAN with a T1 interface, yet the central site has only a single T1 interface. While this configuration allows the deployment to benefit from statistical multiplexing, the router interface at the central site can become congested during traffic bursts, thus degrading voice quality.

- Bursting above Committed Information Rate (CIR)

  Another common configuration is to allow traffic bursts above the CIR, which represents the rate that the service provider has guaranteed to transport across its network with no loss and low delay. For example, a remote site with a T1 interface may have a CIR of only 64 kbps. When more than 64 kbps worth of traffic is sent across the WAN, the provider marks the additional traffic as “discard eligible.” If congestion occurs in the provider network, this traffic will be dropped, possibly having a negative effect on voice quality.

Traffic shaping provides a solution to these issues by limiting the traffic sent out an interface to a rate lower than the line rate, thus ensuring that no congestion occurs on either end of the WAN.

Voice Gateways

This chapter explains important factors to consider when selecting a Cisco voice gateway to provide the appropriate protocol and feature support for your IP Telephony network. The main topics discussed in this chapter include:

Gateway Selection

When selecting an IP Telephony gateway, consider the common or core requirements as well as site and implementation-specific features. Gateways used in IP Telephony applications must meet the following core requirements:

- Dual tone multifrequency (DTMF) relay capabilities

  DTMF relay capability, specifically out-of-band DTMF, separates DTMF digits from the voice stream and sends them as signaling indications through the gateway protocol (H.323, SCCP, or MGCP) signaling channel instead of as part of the voice stream or bearer traffic. Out-of-band DTMF is required when using a low bit-rate codec for voice compression because the potential exists for DTMF signal loss or distortion.

- Supplementary services support

  Supplementary services are typically basic telephony functions such as hold, transfer, and conferencing.

- Fax/modem support
- Cisco CallManager redundancy support

  Cisco Architecture for Voice, Video, and Integrated Data (AVVID) for IP Telephony is based on a distributed model for high availability. Cisco CallManager clusters provide for Cisco CallManager redundancy. The gateways must support the ability to “re-home” to a secondary Cisco CallManager in the event that a primary Cisco CallManager fails. Redundancy differs from call survivability in the event of a Cisco CallManager or network failure. See Call Survivability with Cisco CallManager, page 3-4.

Refer to the gateway product documentation to verify that any IP Telephony gateway you select for an enterprise deployment can support the preceding core requirements. Additionally, every IP Telephony implementation has its own site-specific feature requirements, such as analog or digital access, DID, and capacity requirements (see Site-Specific Gateway Requirements, page 3-5).

Gateway Protocols

Cisco CallManager (Release 3.1 and later) supports the following gateway protocols:

- H.323
- Media Gateway Control Protocol (MGCP)

Cisco IP Phones use Skinny Client Control Protocol (SCCP), which is a lighter-weight protocol. SCCP uses a master/slave model, while H.323 is a peer-to-peer model. MGCP also follows a master/slave model.

Table 3-1 shows which gateways support a given protocol. Each of these protocols follows a slightly different methodology to provide support for the core gateway requirements. For more information on core gateway requirements and protocols, refer to the gateway selection chapter in the *Cisco IP Telephony Solution Reference Network Design* (SRND) for Cisco CallManager Releases 3.1 and 3.2, available online at

http://cisco.com/go/srnd

Table 3-1 Supported Gateway Protocols and Cisco IP Telephony Gateways

Cisco Gateway	MGCP 0.1	H.323	SCCP
VG200[1]	Yes Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes	Yes (DSP farm)
VG248	No	No	Yes[2]
DE-30+, DT-24+[3]	Yes	No	No
Cisco 827-V4	No	Yes, supported for FXS	No
Cisco ATA186/188	Yes, FXS only	Yes, FXS only	Yes, FXS only
Cisco 1751/1760	No	Yes	No
Cisco 2600XM[4]	Yes Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes	DSP farm in Cisco IOS Release 12.2.13T
Cisco 3640 and 3660	Yes Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes	DSP farm in Cisco IOS Release 12.2.13T

Table 3-1 Supported Gateway Protocols and Cisco IP Telephony Gateways (continued)

Cisco Gateway	MGCP 0.1	H.323	SCCP
Cisco 3700[5]	Yes Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes	DSP farm in Cisco IOS Release 12.2.13T
Cisco 5300	No	Yes	No
Cisco AS5350 Cisco AS5400	No	Yes	No
Cisco AS5850	No	Future support	No
Cisco 7200	No	Yes	No
Cisco 7500	No	Yes	No
Catalyst 4000 WS-X4604-GWY Gateway Module	Yes	Yes	No
Catalyst 6000 WS-X6608-x1 Gateway Module and FXS Module WS-X6624	Yes Supported with: • T1 CAS FXS • T1/E1 PRI • FXS with WS-6624	No	No
Cisco ICS7750-MRP	No	Yes	No
Cisco ICS7750-ASI	No	Yes	No
Cisco IAD 2400 Series	Yes	No	No

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.
2. The VG248 is not a true gateway in that it uses Skinny Client Control Protocol (SCCP) instead of H.323 or MGCP.
3. These models have reached end of life.
4. For IP Telephony applications, use Cisco 2600XM routers. For memory considerations for the Cisco 2600 routers, see the Product Bulletin at http://www.cisco.com/warp/customer/cc/pd/rt/2600/prodlit/1675_pp.htm
5. For IP Telephony applications, the Cisco 3700 Series routers provide better performance than the Cisco 3600 Series.

Note Prior to deployment, check the Cisco IOS software release notes to confirm feature or interface support.

Call Survivability with Cisco CallManager

This section describes some general rules for call survivability using H.323, SCCP, and MGCP gateways. Call survivability means that the RTP bearer stream (the voice conversation) between two IP endpoints is preserved even if the Cisco CallManager that originally established the stream between the endpoints is no longer reachable. Endpoints can be categorized as survivable or non-survivable, as indicated in Table 3-2.

Table 3-2 Endpoint Survivability

Survivable Endpoints	Non-Survivable Endpoints
IP Phones (SCCP)	H.323 gateways
MGCP gateways	H.323 endpoints or trunks
	TAPI endpoints (IP SoftPhone, IVR, AutoAttendant, and other Cisco applications)

The following types of failures can occur:

- Cisco CallManager failure — Failures that render Cisco CallManager incapable of responding to call signalling
- Network failure — Any event in the network that prevents communications between Cisco CallManager and the endpoints (for example, gateways or IP phones)

Note Call failure may mean that the phone goes on-hook, receives reorder tone, or in some cases simply goes silent (for example, if the streaming connection is terminated).

Endpoint Rules for Gateway Call Survivability

The following rules apply for various endpoint and Cisco Call Manager failure scenarios:

- If the call involves only non-survivable endpoints, the call fails if there is a failure in any of the Cisco Call Managers to which one of the endpoints is registered. This is true regardless of whether a conference bridge, an MTP, or a transcoder is involved in the call and regardless of which Cisco CallManager (when there are more than one) fails.
- If the call involves one non-survivable endpoint and one survivable endpoint, the call fails only if the Cisco CallManager associated with the non-survivable endpoint fails.
- If the call involves only survivable endpoints and one or more Cisco CallManagers fail, the streaming connection between the endpoints is maintained. However, the endpoints do not have call processing services available to them after the failure. For example, the unavailable services would include transfer, conference, hold, park, pickup, and resume.

In general, MGCP gateways provide the highest degree of call survivability. In Cisco CallManager Release 3.1 and later, the Catalyst 6000 T1/E1 gateway modules use MGCP supporting T1/E1 PRI and T1 CAS, thus enhancing call survivability when an SCCP-based IP phone and an MGCP gateway are the two endpoints. With Cisco IOS Release 12.2.11T and later, IOS-based gateways such as Cisco VG200, 2600XM, 3640, 3660, and 3700 also support MGCP with Cisco CallManager.

Site-Specific Gateway Requirements

Each IP Telephony implementation has its own site-specific requirements. The following questions can help you with IP Telephony gateway selection:

- Is the PSTN (or PBX) access analog or digital?
- What type of analog (FXO, FXS, E&M, DID, CAMA) or digital (T1, E1, CAS, CCS) interface is required for the PSTN or PBX?
- If the PSTN access is digital, what type of signaling is required (T1 CAS, Q.931 PRI, E1 CAS, or R2)?
- What type of signaling does the PBX currently use?
  - FXO or FXS: loop start or ground start
  - E&M: wink-start, delay-start, or immediate-start
  - E&M: type I, II, III, IV, or V
  - T1: CAS, Q.931 PRI (User-Side or Network-Side), QSIG, DPNSS, or Proprietary d-channel (CCS) signaling
  - E1: CAS, R2, Q.931 PRI (User-Side or Network-Side), QSIG, DPNSS, Proprietary d-channel (CCS) signaling
- What type of framing (SF, ESF, or G.704) and line encoding (B8ZS, AMI, CRC-4, or HDB3) does the PBX currently use?
- Does the PBX require passing proprietary signaling? If so, which time slot is the signaling passed on, and is it HDLC-framed?
- What is the required capacity of the gateway; that is, how many channels are required? (Typically, if 12 or more voice channels are required, then digital is more cost effective than an analog solution.)
- Is Direct Inward Dialing (DID) required? If so, specify analog or digital.
- Is Calling Line ID (CLID) needed?
- Is Calling Name needed?
- What types of fax and modem support are required?
- What types of voice compression are required?
- What types of supplementary services are required?
- Will the PBX provide clocking, or will it expect the Cisco gateway to provide clocking?
- Is rack space available for all needed gateways, routers, and switches?

Note Direct Inward Dial (DID) refers to a private branch exchange (PBX) or Centrex feature that permits external calls to be placed directly to a station line without use of an operator.

Note Calling Line Identification (CLI, CLID, or ANI) refers to a service available on digital phone networks to display the calling number to the called party. The central office equipment identifies the phone number of the caller, enabling information about the caller to be sent along with the call itself. CLID is synonymous with Automatic Number Identification (ANI).

Cisco IP Telephony gateways are able to inter-operate with most major PBX vendors, and they are EIA/TIA-464B compliant.

The following tables summarize the features and interface types supported by the various Cisco gateway models.

Note In the following tables, the Cisco IOS and Cisco CallManager release numbers refer to the minimum release that can support the listed feature on a particular gateway platform. For specific recommendations about the preferred software release for each hardware platform, see Recommended Hardware and Software Combinations, page A-1.

Cisco Analog Gateways

Table 3-3 lists supported interface types for Cisco analog gateways using H.323 or Session Initiation Protocol (SIP), and Table 3-4 lists supported interface types for Cisco analog gateways using Media Gateway Control Protocol (MGCP).

Table 3-3 Supported Analog H.323 and SIP Features

	Interface Type					
Cisco Gateway	**FXS**	**FXO**	**E&M**	**FXO, Battery Reversal**	**Analog DID**	**CAMA 911**
Analog Telephone Adapter (ATA)	Yes	No	No	No	No	No
827-4V	Yes	No	No	No	No	No
1751 and 1760	Yes	Yes	Yes	Yes	Yes	Future
IAD2400	No	No	No	No	No	No
VG200[1]	Yes	Yes	Yes	No	Yes	12.2.13T
VG248	No	No	No	No	No	No
2600 Series	Yes	Yes	Yes	Yes	Yes	12.2.11T
3600 Series	Yes	Yes	Yes	Yes	Yes	12.2.11T
3700 Series	12.2.8T	12.2.8T	12.2.8T	12.2.8T	12.2.8T	12.2.11T
ICS 7750	Yes	Yes	Yes	Yes	Yes	Future
WS-6624	No	No	No	No	No	No
Catalyst 6500 Series Communication Media Module (CMM) with FXS module	Future	No	No	No	No	No
7x00 family	No	No	No	No	No	No

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.

Table 3-4 Supported Analog MGCP Features

Cisco Gateway	Interface Type					
	FXS	FXO	E&M	FXO, Battery Reversal	Analog DID	CAMA 911
Analog Telephone Adapter (ATA)	Yes	No	No	No	No	No
827-4V	No	No	No	No	No	No
1751 and 1760	Future	Future	No	No	No	No
IAD2400	Yes	Yes	No	No	No	No
VG200[1]	Yes	Yes	No	No	No	No
VG248	No	No	No	No	No	No
2600 Series	12.2.4T	12.2.4T	No	No	No	No
3600 Series	12.2.4T	12.2.4T	No	No	No	No
3700 Series	12.2.8T	12.2.8T	No	No	No	No
ICS 7750	Yes	Yes	No	No	No	No
WS-6624	Yes	No	No	No	No	No
Catalyst 6500 Series Communication Media Module (CMM) with FXS module	Yes	No	No	No	No	No
7x00 family	No	No	No	No	No	No

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.

Cisco Digital Gateways

Table 3-5 through Table 3-8 list supported interface types for Cisco digital gateways using H.323 or Session Initiation Protocol (SIP). Table 3-9 lists supported interface types for Cisco digital gateways using Media Gateway Control Protocol (MGCP).

Table 3-5 Supported Digital H.323 and SIP Features for BRI, T1 CAS, T1 FGD, and T1 QSIG

Cisco Gateway	Interface Type						
	BRI (TE, User side)	BRI (NT, Network side)	BRI QSIG (Net3)	BRI Phones	T1 CAS (Robbed bit)	T1 FGD	T1 QSIG
Analog Telephone Adapter (ATA)	No	No	No	No	No	No	No
827-4V	No	No	No	No	No	No	No
1751 and 1760	No	Yes	Yes	No	12.2.4YB	No	12.2.4YB
IAD2400	No	No	No	No	No	No	No
VG200[1]	Yes	Yes	No	No	Yes	Yes	No
VG248	No	No	No	No	No	No	No

Table 3-5 Supported Digital H.323 and SIP Features for BRI, T1 CAS, T1 FGD, and T1 QSIG (continued)

Cisco Gateway	Interface Type						
	BRI (TE, User side)	BRI (NT, Network side)	BRI QSIG (Net3)	BRI Phones	T1 CAS (Robbed bit)	T1 FGD	T1 QSIG
2600 Series	Yes	Yes	Yes	No	Yes	Yes	Yes
3600 Series	Yes	Yes	Yes	No	Yes	Yes	Yes
3700 Series	Yes	Yes	Yes	No	Yes	Yes	Yes
ICS 7750	Yes	Yes	Future	No	Yes	Future	Future
WS-6608	No	No	No	No	No	No	No
Catalyst 6500 Series Communication Media Module (CMM) with T1/E1 modules	No	No	No	No	Future	No	Future
7x00 family	No	No	No	No	Yes	Yes	Yes

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.

Table 3-6 Supported Digital H.323 and SIP Features for T1 PRI DMS-100, 4ESS, and 5ESS

Cisco Gateway	Interface Type					
	T1 PRI (User, DMS-100)	T1 PRI (Network, DMS-100)	T1 PRI (User, 4ESS)	T1 PRI (Network, 4ESS)	T1 PRI (User, 5ESS)	T1 PRI (Network, 5ESS)
Analog Telephone Adapter (ATA)	No	No	No	No	No	No
827-4V	No	No	No	No	No	No
1751 and 1760	12.2.4YB	Future	12.2.4YB	Future	12.2.4YB	Future
IAD2400	No	No	No	No	No	No
VG200[1]	Yes	No	Yes	No	Yes	No
VG248	No	No	No	No	No	No
2600 Series	Yes	Future	Yes	Future	Yes	Future
3600 Series	Yes	Future	Yes	Future	Yes	Future
3700 Series	Yes	Future	Yes	Future	Yes	Future
ICS 7750	Yes	Future	Yes	Future	Yes	Future
WS-6608	No	No	No	No	No	No
Catalyst 6500 Series Communication Media Module (CMM) with T1/E1 modules	Future	Future	Future	Future	Future	Future
7x00 family	Yes	Future	Yes	Future	Yes	Future

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.

Table 3-7 Supported Digital H.323 and SIP Features for T1 PRI NI2, NFAS, and Network Specific Facilities (NSF) Service

	Interface Type					
Cisco Gateway	**T1 PRI (User, NI2)**	**T1 PRI (Network, NI2)**	**T1 PRI NFAS (User, DMS-100)**	**T1 PRI NFAS (User, 4ESS)**	**T1 PRI NFAS (User, 5ESS)**	**T1 PRI (Megacom or SDN, 4ESS)**
Analog Telephone Adapter (ATA)	No	No	No	No	No	No
827-4V	No	No	No	No	No	No
1751 and 1760	12.2.4YB	12.2.4YB	No	No	No	No
IAD2400	No	No	No	No	No	No
VG200[1]	Yes	Yes	No	No	No	No
VG248	No	No	No	No	No	No
2600 Series	Yes	Yes	Yes	Yes	Yes	Yes
3600 Series	Yes	Yes	Yes	Yes	Yes	Yes
3700 Series	Yes	Yes	Yes	Yes	Yes	Yes
ICS 7750	Yes	Yes	Yes	Yes	Yes	No
WS-6608	No	No	No	No	No	No
Catalyst 6500 Series Communication Media Module (CMM) with T1/E1 modules	Future	Future	Future	Future	Future	Future
7x00 family	Yes	Yes	No	No	No	No

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.

Table 3-8 Supported Digital H.323 and SIP Features for E1 and J1

	Interface Type						
Cisco Gateway	**E1 CAS**	**E1 MELCAS**	**E1 R2**	**E1 PRI (User side, Net5)**	**E1 PRI (Network side, Net5)**	**E1 QSIG**	**J1**
Analog Telephone Adapter (ATA)	No	No	No	No	No	No	No
827-4V	No	No	No	No	No	No	No
1751 and 1760	No	No	12.2.4YB	12.2.4YB	12.2.4YB	12.2.4YB	No
IAD2400	No	No	No	No	No	No	No
VG200[1]	No	Yes	Yes	Yes	Yes	No	12.2.13T
VG248	No	No	No	No	No	No	No

Table 3-8 Supported Digital H.323 and SIP Features for E1 and J1 (continued)

	Interface Type						
Cisco Gateway	**E1 CAS**	**E1 MELCAS**	**E1 R2**	**E1 PRI (User side, Net5)**	**E1 PRI (Network side, Net5)**	**E1 QSIG**	**J1**
2600 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3600 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3700 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ICS 7750	No	No	No	Yes	Yes	No	No
WS-6608	No	No	No	No	No	No	No
Catalyst 6500 Series Communication Media Module (CMM) with T1/E1 modules	Future	Future	Future	Future	Future	Future	No
7x00 family	Yes	No	Yes	Yes	Yes	Yes	No

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.

Table 3-9 Supported Digital MGCP Features

	Interface Type					
Cisco Gateway	**BRI**	**T1 CAS (E&M)**	**T1 PRI**	**T1 QSIG**	**E1 PRI**	**E1 QSIG**
Analog Telephone Adapter (ATA)	No	No	No	No	No	No
827-4V	No	No	No	No	No	No
1751 and 1760	No	Future	Future	Future	Future	Future
IAD2400	No	Yes	No	No	No	No
VG200[1]	No	Yes	Yes	CallManager 3.3	Yes	CallManager 3.3
VG248	No	No	No	No	No	No
2600 Series	No	Yes[2]	Yes[2]	CallManager 3.3	Yes[2]	CallManager 3.3
3600 Series	No	Yes[2]	Yes[2]	CallManager 3.3	Yes[2]	CallManager 3.3
3700 Series	No	Yes[2]	Yes[2]	CallManager 3.3	Yes[2]	CallManager 3.3
ICS 7750	No	Yes	Yes	CallManager 3.3	Yes	CallManager 3.3
WS-6608	No	Yes	Yes	CallManager 3.3	Yes	CallManager 3.3
Catalyst 6500 Series Communication Media Module (CMM) with T1/E1 modules	No	Yes	Yes	CallManager 3.3	Yes	CallManager 3.3
7x00 family	No	No	No	No	No	No

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.
2. AIM-VOICE-30 modules will support MGCP in Cisco IOS Release 12.2.13T.

QSIG Support

QSIG is a suite of international standards designed to provide flexibility in connecting PBX equipment to a corporate network. Among its other features, QSIG provides an open, standards-based method for interconnecting PBX equipment from different vendors. Detailing the benefits of QSIG is beyond the scope of this document, but for more information, refer to the QSIG documentation at

http://www.qsig.ie/

ECMA QSIG is currently supported in H.323 gateways in PBX-to-PBX mode. The H.323 gateways provide full QSIG feature transparency for QSIG information elements. Basic call setup and teardown are supported using H.323 QSIG gateways, as summarized in Table 3-10.

Table 3-10 QSIG Support on H.323 Gateways

Platform	Media	Minimum Cisco IOS Software Release Required
Cisco 2600 and 3600 Series	BRI and T1/E1 QSIG	12.1.2T
Cisco 3810	BRI and T1/E1 QSIG	12.0.4T
Cisco 7200	T1/E1 QSIG	12.1.2T
Cisco 7500	T1/E1 QSIG	12.1.3T
Cisco 5300	T1/E1	12.0.7T
Cisco AS5350 Cisco AS5400	T1/E1	12.2.2T

For more information on QSIG support on Cisco IOS gateways, refer to

http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121newft/121t/121t2/dt_qsig.htm#xtocid116542

Prior to Cisco CallManager Release 3.3, basic PRI functionality is all that is provided whenever a PBX is connected to a gateway using QSIG via H.323 and calls are made between phones on the PBX and IP phones attached to the Cisco CallManager. This basic functionality, which includes only the Calling Line Identifier (CLID) and Direct Inward Dialed (DID) number, is provided by the gateway terminating the QSIG protocol rather than by Cisco CallManager.

For Cisco CallManager to support QSIG functionality, QSIG must be back-hauled directly to Cisco CallManager. This support is provided in Cisco CallManager Release 3.3 and later, in conjunction with MGCP gateways such as the Catalyst 6608, VG200, 2600XM Series, and 3640/60 Series.

Fax and Modem Support

This section describes the fax and modem support available with Cisco CallManager and Cisco voice gateways. This section first presents brief overviews of fax and modem support on Cisco voice gateways, followed by a listing of supported platforms and example configuration files.

Gateway Support for Fax Pass-Through and Cisco Fax Relay

Fax over IP enables interoperability of traditional analog fax machines with IP Telephony networks. The fax image is converted from an analog signal and is carried as digital data over the packet network.

In its original form, fax data is digital. However, to transmit across a traditional PSTN, it is modulated and converted to analog. Fax over IP reverses this analog conversion, transmitting digital data over the packet network and then reconverting the digital data to analog for the receiving fax machine.

Most Cisco voice gateways currently support two methods to transmit fax traffic across the IP network:

- Cisco Fax Relay — In fax relay mode, the gateways terminate the T.30 fax signaling.
- Fax Pass-Through — In fax pass-through mode, the gateways do not distinguish a fax call from a voice call.

Cisco Fax Relay mode is the preferred method to transmit fax traffic. However, if a specific gateway does not support Cisco fax relay, it supports fax pass-through.

Best Practices

The following recommendations and guidelines can assist you in best implementing fax support on Cisco voice gateways:

- When using QoS, make every effort to minimize the following:
  - Packet loss
  - Delay
  - Delay variation (jitter)

  For detailed information about implementing QoS in a Cisco IP Telephony network, refer to the *Cisco AVVID Network Infrastructure Enterprise Quality of Service Design* guide, available at

  http://cisco.com/go/srnd
- The following tips can help ensure the integrity of the fax calls:
  - Use call admission control (CAC) to ensure that calls are not admitted if they exceed the specified total bandwidth limit.
  - Disable call waiting on all dedicated modem and fax ports.
- For best performance, verify that you have Cisco Fax Relay on both the originating and terminating gateways. If two Cisco IOS gateways have differing transports, they will negotiate to use Cisco Fax Relay.

  The only non-IOS gateway that does not support Cisco Fax Relay is the Cisco Digital Access DT-24/DE-30+. If you connect this gateway to a Cisco IOS gateway, you should configure both gateways to use fax pass-through mode.
- Ensure that constant packet delay on the network does not exceed 1 second and that delay variation (jitter) does not exceed 240 milliseconds.

- To improve performance in networks with a high frequency of out-of-order packet arrival, disable Error Correction Mode (ECM) on the fax machines.
- Most fax machines appear to accept packet drop in the range of 0.4% to 0.6% without slowing down to the next speed. However, in a network with packet drop in the range of 0.8% to 1%, you should disable ECM.
- You can disable ECM on the gateway itself rather than disabling it on multiple fax machines. However, if packet drops occur, the fax image quality might deteriorate. Therefore, you should disable ECM only after considering whether you want to risk compromising image quality rather than experiencing longer call durations or dropped calls. You should also monitor and evaluate the network to identify and resolve the cause of the dropped packets.

Gateway Support for Modem Pass-Through

In general, there are two mechanisms for supporting modem sessions over an IP network using voice gateways:

- Modem pass-through
- Modem relay

Currently, modem pass-through is the only mechanism supported on Cisco voice gateways.

Modem pass-through is the transport of modem signals through a packet network using pulse code modulation (PCM) encoded packets and a G.711 codec. Modem pass-through requires the ability of the gateways to discriminate between modem signals and voice signals and take appropriate action. When the gateway detects the modem signal, it disables the following services:

- Echo cancellation (EC)
- Voice activity detection (VAD)

In modem pass-through mode, the gateways do not distinguish a modem call from a voice call. The communication between the two modems is carried in-band in its entirety over a "voice" call. The modem traffic is transparently carried over a QoS-enabled IP infrastructure, and at no point is the data demodulated within the IP network.

Modem upspeed is similar to pass-through in the sense that the modem call is carried in-band over the "voice" call. The difference is that the gateways are, to some extent, aware of the modem call when the upspeed feature is used. Although relay mechanisms are not employed, the gateways do recognize the modem tone, automatically change the "voice" codec to G.711 (the "upspeed" portion), and turn off VAD and echo cancellation (EC) for the duration of the call.

Currently, this upspeed feature is not supported on any Cisco IOS platform except the Cisco AS5300 via Cisco IOS Release 12.1.3T. For Cisco 2600, 3600, VG200, 4224, and Catalyst 4000 Access Gateway Module (AGM) platforms, the modem upspeed feature will be supported in a future Cisco IOS release. For these platforms, you can configure **no vad** on the dial peer until the modem upspeed feature becomes available.

The modem upspeed feature is also supported on the Catalyst 6000 gateway modules.

Best Practices

Observe the following recommended best practices to ensure optimum performance of modem traffic transported over an IP infrastructure:

- Ensure that the IP network is enable for Quality of Service (QoS) and that you adhere to all of the recommendations for providing QoS in the LAN, MAN, and WAN environments. Every effort should be made to minimize the following parameters:
  - Packet loss — Fax and modem traffic requires an essentially loss-free transport. A single lost packet will result in retransmissions.
  - Delay
  - Delay variation (jitter)

  For more information, refer to the *Cisco AVVID Network Infrastructure Enterprise Quality of Service Design* guide, available at

  http://cisco.com/go/srnd
- Use call admission control (CAC) to ensure that calls are not admitted if they exceed the specified total bandwidth limit.
- Use G.711 for all calls involving a modem. If one of the gateways does not support modem relay, modem pass-through will be negotiated (G.711 only). If modems are used, the best-practice recommendation is to use G.711 for all calls.
- Do not use the IP network to connect modems that will be used to troubleshoot or diagnose problems with the IP network. In this case, the modems used to troubleshoot the devices that compose the IP infrastructure should be connected to a plain old telephone service (POTS).
- Where possible, use a single signaling protocol and gateway family to minimize interoperability issues.
- Disable call waiting on all dedicated modem and fax ports.

V.90 Support

Currently, Cisco equipment supports only V.34 modems. Although V.90 modems will function on existing hardware, and speeds higher than V.34 speeds can be achieved, full V.90 support cannot be guaranteed.

Supported Platforms and Features

The following Cisco platforms support fax and modem features:

Analog Gateways

Cisco IOS Gateways:

- 2600XM (VIC-2FXS)
- VG200
- 4224
- AGM
- 1760V
- 1751
- 827-4V

Non-IOS Gateways:

- VG248
- ATA186

Digital Gateways

Cisco IOS Gateways:

- 3600 (NM-HDV)
- 3700
- 7200
- 7500
- AS5300
- AS5350
- AS5400

Non-IOS Gateways:

- 6608
- 6624

Note Fax and modem support was tested on the above platforms using Cisco IOS Release 12.3(1) on the Cisco IOS gateways and Release 1.2.1 of the Cisco VG248 Analog Phone Gateway.

Platform Protocol Support

Common call control protocols used today in enterprise solutions include H.323, Media Gateway Control Protocol (MGCP), and Skinny Client Control Protocol (SCCP). Not all Cisco voice platforms support all of these protocols or all of the fax and modem features, thus raising interoperability issues. Additional interoperability issues occur when mixing Cisco IOS gateways, such as the Cisco 2600XM or the Cisco 3600 Series, with non-IOS gateways such as the VG248. This section lists the combinations of gateways that provide support for interoperability of fax, modem, and protocol features.

At a high level, Cisco IOS Release 12.3(1) – Load 47 on the Cisco 6608 and Load 41 on the Cisco 6624 – and Release 1.2.1 on the VG248 do support interoperability of Cisco fax relay, modem pass-through, and voice features. Prior to Cisco IOS Release 12.2(11)T1, only voice and Cisco fax relay were supported between Cisco IOS and non-IOS voice platforms because incompatibility of the pass-through Network Services Engine (NSE) scheme prevented modem pass-through from interoperating.

Some of the common combinations of protocols in a network include MGCP and H.323, SCCP and H.323, and SCCP and MGCP. Common voice gateways included the Cisco VG200, VG248, 2600XM, 3600, 5300, and Catalyst 6000.

Table 3-11 lists the protocol combinations that currently support fax and modem interoperability.

Table 3-11 Fax and Modem Features Supported with Various Combinations of Call Control Protocols

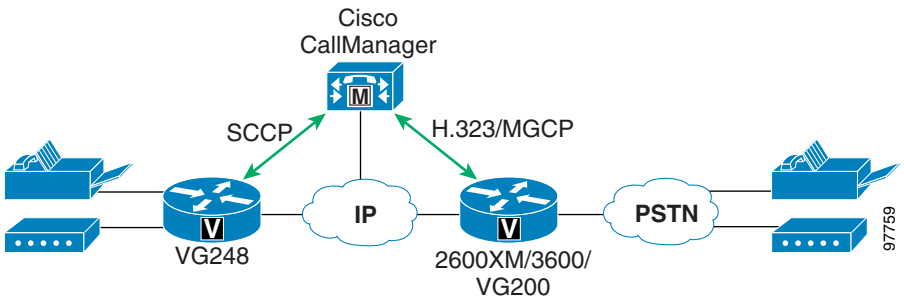
Protocol Combinations	Modem Relay	Modem Pass-Through	T.38 Fax Relay	Cisco Fax Relay	Fax Pass-Through
Cisco CallManager using MGCP combined with Cisco CallManager using H.323	Yes	Yes	Yes	Yes	Yes
Cisco CallManager using MGCP combined with Cisco CallManager using MGCP	Yes	Yes	Yes	Yes	Yes
SCCP combined with Cisco CallManager using H.323	Yes	Yes	Yes	Yes	Yes
SCCP combined with Cisco CallManager using MGCP	Yes	Yes	Yes	Yes	Yes

Note Cisco ATA186, VG248 and Catalyst 6000 platforms currently do not support T.38 fax relay. When these platforms connect to Cisco AS5350 or AS5400 gateways, only fax pass-through is supported for fax applications.

Gateway Combinations and Interoperability of Features

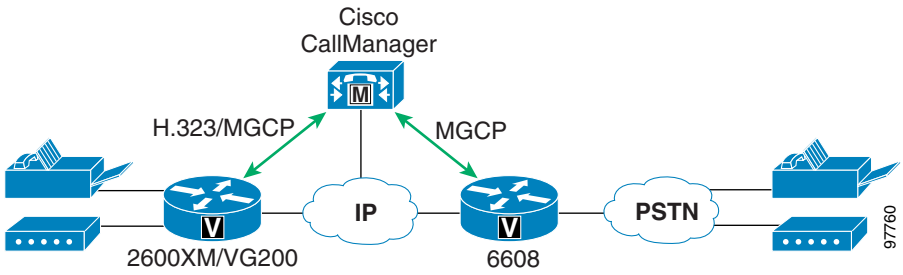
The most frequent questions about fax and modem interoperability arise from combining a Cisco IOS gateway (such as a Cisco 2600XM or 3600) with a non-IOS gateway (such as a Cisco VG248), as illustrated in Figure 3-1.

Figure 3-1 Configuration Combining a Cisco IOS and Non-IOS Gateway



The second most common source of questions about fax and modem interoperability arise in configurations using only Cisco IOS gateways, as illustrated in Figure 3-2.

Figure 3-2 Configuration Using Only Cisco IOS Gateways



The answer is basically the same for both scenarios: Prior to Cisco IOS Load 47 on the 6608 and Release 1.2.1 on the VG248, only voice and Cisco fax relay are supported, while fax and modem pass-through are not supported because of NSE incompatibility. With Cisco IOS Load 47 or later on the 6608, Load 41 or later on the 6624, and Release 1.2.1 on the VG248, all three platforms can interoperate with Cisco IOS gateways for voice, Cisco fax relay, and modem pass-through, regardless of call control protocol. The NSE pass-through scheme is independent of call control protocol because it operates in the bearer path instead of the signaling path.

Feature Support Between Similar Gateways

Table 3-12 lists the fax and modem features supported between gateways of the same general type, such as between the Cisco VG248 and 6608, between 2600XM and 3600, or between 2600XM and AS5300. In these cases, as long as both platforms support a given feature, those platforms will interoperate.

Table 3-12 Fax and Modem Feature Support on Gateways of the Same Type

Gateway Type	Fax Pass-Through	Cisco Fax Relay	T.38 Fax Relay	Modem Pass-Through	Modem Relay
Cisco IOS gateways	Supported	Supported (except on 5350 and 5400)	Supported	Supported	Supported (only on NM-HDV)
Non-IOS gateways	Supported	Supported (except on ATA186)	N/A	Supported (except on ATA186)	N/A

Gateway Configuration Examples

This section provides listings of example gateway configurations for fax and modem support.

Cisco IOS Gateway Configuration

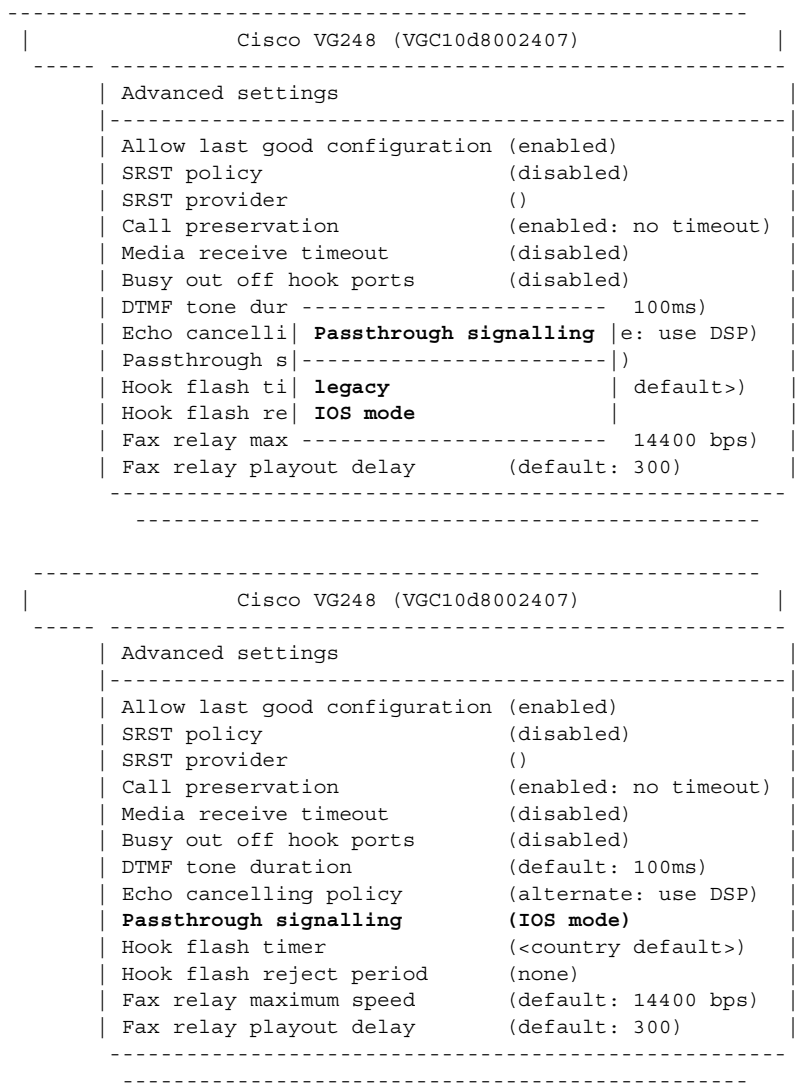
H.323

```
!
! Cisco fax relay is ON by default
!(except for 5350/5400, where Cisco fax relay is not supported)
!
dial-peer voice 1000 voip
 destination-pattern 1T
 session target ipv4:10.10.10.1
 modem passthrough mode nse codec g711ulaw
!
!
```

MGCP

```
!
ccm-manage mgcp
mgcp
mgcp call-agent 10.10.10.1 service-type mgcp version 0.1
mgcp modem passthrough voip mode nse
mgcp fax t38 inhibit
!
dial-peer voice 100 pots
 application mgcpapp
 port 1/0/0
!
```

Cisco VG248 Configuration



Cisco CallManager Configuration for Cisco IOS Gateways

Perform the following steps in Cisco CallManager to configure it for the Cisco IOS gateways (such as the Cisco 6608 and 6624). Also see Figure 3-3 and Figure 3-4 for an illustration of these configuration steps.

Step 1 In Cisco CallManager Administration, select **Device > Gateway** to display the **Find/List Gateways** window.

Step 2 Search for the gateway you want to modify (if it already exists), or click on **Add a New Gateway** to add a new gateway to the Cisco CallManager database.

Step 3 After selecting the appropriate type of gateway (for example, Cisco Catalyst 6000), click on **Fax Relay Enable** to enable Cisco fax relay.

Step 4 Using the **NSE Type** drop-down list box, select **IOS Gateways** for modem pass-through.

Step 5 Click **Update** to save the changes.

Step 6 Reset the gateway to apply the changes.

Figure 3-3 Gateway Configuration in Cisco CallManager

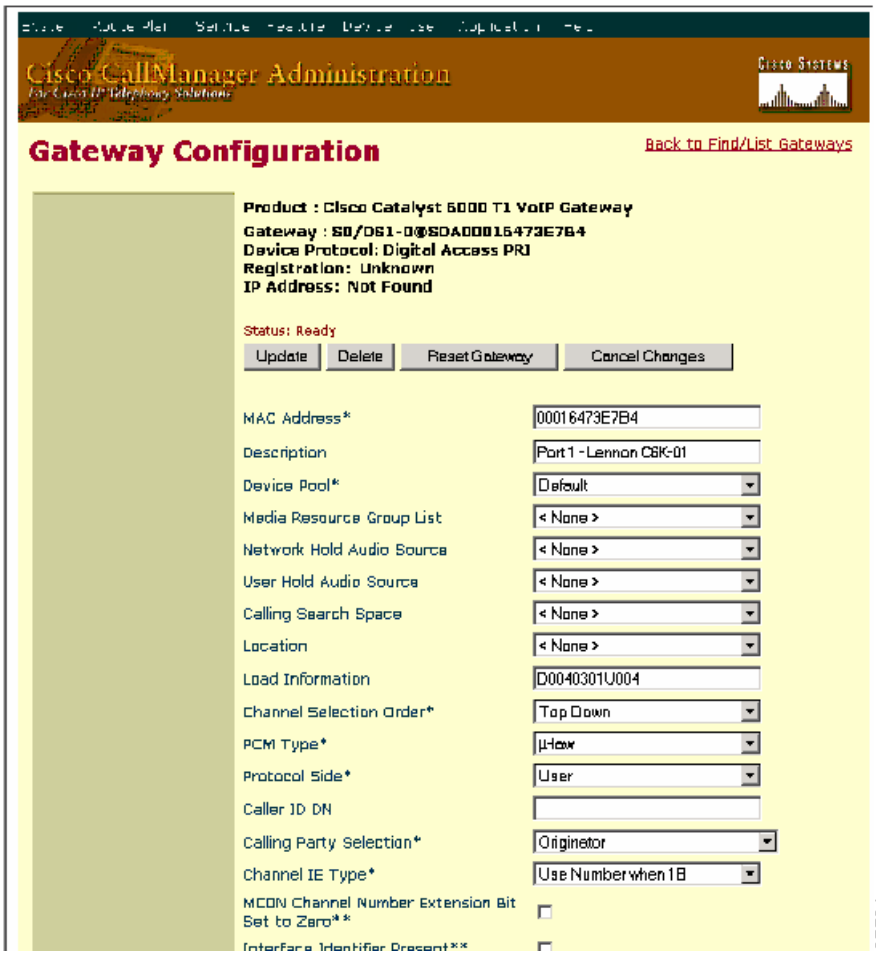
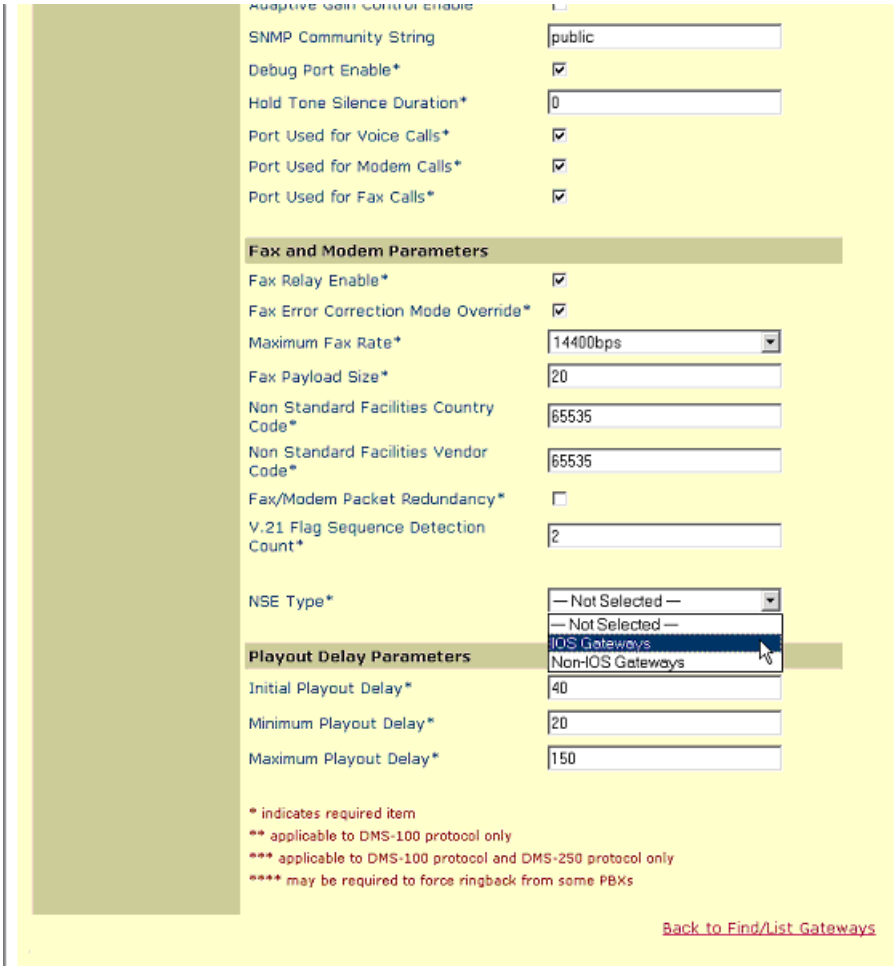


Figure 3-4 Gateway Configuration in Cisco CallManager (continued)



This configuration supports voice, Cisco fax relay, and modem pass-through between Cisco VG248, 6608, 6624, and IOS gateways, with the exception of Cisco AS5350 and AS540 gateways (which do not support Cisco fax relay). The configuration also supports a V.34 modem connection in pass-through mode. V.90 modem connections are not guaranteed but are possible, depending on amount of network jitter and clock sync.

Clock Sourcing for Fax and Modem Pass-Through

The clock signal plays a critical role in enabling fax and modem pass-through to work correctly. The gateway clock must synchronize with the PSTN clock, where Stratum clocking is provided. Without this clock synchronization, fax and (especially) modem pass-through will not work. To synchronize the clocks correctly, enter the following configuration in the T1 controller. (In this example, the T1 controller is the voice gateway that connects to the PSTN.)

```
!
controller T1 0
 framing esf
 linecode b8zs
 clock source line
 channel-group 1 timeslots 1-24 speed 64
 !
```

Also enter this configuration in all other interfaces connected to the PSTN.

T.38 Fax Relay

T.38 fax relay is not supported on Cisco ATA186, VG248, 6608, and 6624 gateways, but it is supported on most of the high-performance Cisco IOS voice platforms such as the Cisco 2600XM and 3600.

You can configure T.38 fax relay in any of the following ways:

- Loose Gateway Controlled with Network Services Engine (NSE), page 3-21
- Gateway Controlled with Capability Exchange Through H.245 or Session Definition Protocol (SDP), page 3-22
- Call-Agent-Controlled T.38 with H.323 Annex D and MGCP, page 3-23

Loose Gateway Controlled with Network Services Engine (NSE)

This configuration uses static T.38 configuration on the dial-peer, as illustrated in the following Cisco IOS gateway configuration example:

H.323

```
!
dial-peer voice 1000 voip
 destination-pattern 1T
 session target ipv4:10.10.10.1
 modem passthrough mode nse codec g711ulaw
 fax protocol t38
!
!
```

MGCP

```
!
ccm-manage mgcp
mgcp
mgcp call-agent 10.10.10.1 service-type mgcp version 0.1
mgcp modem passthrough voip mode nse
no mgcp fax t38 inhibit
!
```

```
dial-peer voice 100 pots
 application mgcpapp
 port 1/0/0
!
!
```

Gateway Controlled with Capability Exchange Through H.245 or Session Definition Protocol (SDP)

The following characteristics apply to this method of configuring T.38 fax relay:

- T.38 capability is exchanged between gateways. A Network Services Engine (NSE) message is sent on the RTP stream from the terminating gateway to signal the originating gateway to switch to T.38 fax relay upon detection of a fax tone. Because the NSE message is sent on the RTP stream, it is transparent to call control signaling.
- Cisco CallManager cannot support this capability exchange with either H.323 or MGCP. Therefore, you must use a configuration command to force T.38 fax relay even though T.38 capability is not exchanged.
- There are three fallback mechanism to choose from:
  - Cisco fax relay (default)
  - Fax pass-through
  - None

The following example illustrates this type of configuration:

H.323

```
!
dial-peer voice 1000 voip
 destination-pattern 1T
 session target ipv4:10.10.10.1
 modem passthrough mode nse codec g711ulaw
!
! To enable T.38 fax relay and fall back to Cisco fax relay when
! T.38 fax negotiation fails. This is the default case.
 fax protocol t38 fallback cisco
!
dial-peer voice 1001 voip
 destination-pattern 2T
 session target ipv4:10.10.10.2
 modem passthrough mode nse codec g711ulaw
!
! To enable T.38 fax relay and fall back to fax passthrough when
! T.38 fax negotiation fails.
fax protocol t38 nse fallback pass-through
!
dial-peer voice 1002 voip
 destination-pattern 3T
 session target ipv4:10.10.10.3
 modem passthrough mode nse codec g711ulaw
!
! This CLI is needed when talking to MGCP endpoint where CA/GK
! doesn't support T.38 fax relay such as CCM.
 fax protocol t38 nse force fallback none
!
!
```

MGCP

```
!
ccm-manage mgcp
mgcp
mgcp call-agent 10.10.10.1 service-type mgcp version 0.1
mgcp modem passthrough voip mode nse
no mgcp fax t38 inhibit
!
! This CLI is needed when CA doesn't support T.38 fax relay
mgcp fax t38 gateway force
!
dial-peer voice 100 pots
 application mgcpapp
 port 1/0/0
!
!
```

In topologies that employ the Cisco VG248 and 6608 or 6624, use the following Cisco IOS commands:

```
fax protocol t38 [nse [force]] fallback [cisco | none]
modem passthrough nse codec {g711ulaw|g711alaw}
```

These two commands enable Cisco IOS gateways to interoperate with the VG248 for Cisco fax relay and modem pass-through as well as with other Cisco IOS gateways for T.38 fax relay and modem pass-through.

Call-Agent-Controlled T.38 with H.323 Annex D and MGCP

The following characteristics apply to this method of configuring T.38 fax relay:

- The call control agent (for example, Cisco CallManager) controls the T.38 fax relay, and the gateways operate in passive mode.
- No NSE messages are sent from gateway to gateway.
- In this type of configuration, the T.38 fax relay is *not* transparent to the call control protocol. The call agent performs the protocol translation between H.323 and MGCP.
- This method of configuring T.38 fax relay is available with Cisco IOS Release 12.3(1). The Cisco BTS 10200 Softswitch also supports this method.
- Cisco CallManager does not support call agent control of T.38 fax relay. Therefore, this method of configuring T.38 fax relay does *not* apply to Cisco CallManager deployments.

The following example illustrates this type of configuration:

H.323

```
!
dial-peer voice 1000 voip
 destination-pattern 1T
 session target ipv4:10.10.10.1
 modem passthrough mode nse codec g711ulaw
!
! To enable T.38 fax relay.
 fax protocol t38
!
!
```

MGCP

```
!
ccm-manage mgcp
mgcp
mgcp call-agent 10.10.10.1 service-type mgcp version 0.1
!
! T.38 fax relay is ON by default. HOWEVER, CCM doesn't
! support CA controlled mode. This is the configuration for
! talking to BTS.
!
dial-peer voice 100 pots
 application mgcpapp
 port 1/0/0
!
```

Media Resources

A media resource is a software-based 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 circuit, packetization of a stream, and so forth.

This chapter focuses on the following high-level functions of media resources:

- Media termination point (MTP)
- Voice termination
- Conferencing
- Transcoding

Media Resource Hardware

Hardware-based media resources use digital signal processors (DSPs) to perform the media processing functions. Table 4-1 lists the Cisco hardware platforms that can provide media resources, along with the number and type of DSPs available on each platform.

Table 4-1 DSP Resources Available on Each Hardware Platform

Cisco Hardware Platform	Hardware Configuration	Number of DSPs	Type of DSP Chipset	Usage
NM-HD-2VE	Fixed	3	TI 5510	Voice termination
NM-HD-2V	Fixed	1	TI 5510	Voice termination
NM-HD-1V	Fixed	1	TI 5510	Voice termination
NM-HDV or NM-HDV-FARM	Upgradeable: • 1-5 packet voice/data modules (PVDMs) • Each PVDM has 3 DSPs	3, 6, 9, 12, or 15	TI 549	Voice termination, conferencing, and transcoding[1]
NM-HDA	Upgradeable	2 or 4	TI 5421	Voice termination
AIM-VOICE-30	Fixed	4	TI 5421	Voice termination

Table 4-1 DSP Resources Available on Each Hardware Platform (continued)

Cisco Hardware Platform	Hardware Configuration	Number of DSPs	Type of DSP Chipset	Usage
Catalyst 6000 WS-6608-T1 or WS-6608-E1	Fixed	64	TI 549	Voice termination, conferencing, and transcoding
Catalyst 4000 WS-X4604-GWY	Fixed	24	TI 549	Voice termination, conferencing, and transcoding
ICS 7750 MRP with WAN interface cards (WICs)	Fixed	7, 9, or 10[2]	TI 549	Voice termination, transcoding
VG248	Fixed	24	TI 549	Voice termination
WS-6624-FXS	Fixed	12	TI 549	Voice termination

1. Conferencing and transcoding on are supported after Cisco IOS Release 12.2(13)T.
2. Number of DSPs depends on the specific module.

Voice Termination

The number of voice terminations that a platform can support depends on the chipset used by that platform. The number of voice terminations that a DSP can support also depends on the complexity of the codec being used. *Codec complexity* is a method of grouping codecs according to their processing overhead.

You can assign a complexity level to a codec based on its processing overhead to the DSP. High complexity codecs can support fewer calls than medium complexity codecs. If no more DSP resources of the required codec type are available to support a new call, that call will be rejected. Therefore, before assign complexity levels, you should carefully consider how many calls of each codec type the DSPs will have to support.

Cisco hardware platforms currently use three models of DSPs: the TI 549, TI 5421, and TI 5510.

TI 549 and TI 5421

The TI 549 and TI 5421 DSPs support only medium complexity and high complexity codecs. When using these DSPs, you must define all codecs as one of these two complexities. The TI 5421 essentially combines two TI 549s in a single package. Table 4-2 lists all the codec types and the corresponding number of channels supported on the TI 549 and TI 5421.

Table 4-2 Codec Channels Supported on TI 549 and TI 5421 DSPs

Codec Type[1]	TI 549 (Channels per DSP)	TI 5421 (Channels per DSP)
Medium Complexity: 711 (A-law, mu-law) G.726 (16K, 24K, 32K) GSM Full Rate (GSMFR) G.729a G.729ab Fax Relay Clear-channel	4	8
High Complexity: G.723.1 (5.3K, 6.3K) G.723.1.a (5.3K, 6.3K) G.728 GSM Enhanced Full Rate (GSMEFR) G.729 G.729b	2	4

1. Not all hardware platforms support all codec types or DSP types. See Table 4-1 for DSP and codec types supported on each platform.

TI 5510

The TI 5510 supports medium and high complexity, and also supports flex mode. Flex mode removes the requirement to define the codec complexity at configuration time, but the trade-off is that flex mode also allows oversubscription of DSP resources.

On the TI 5510, all G.729 codecs are defined as medium complexity. Table 4-3 lists all the codec types and the corresponding number of channels supported on the TI 5510.

Table 4-3 Codec Channels Supported on the Cisco NM-HD-xx (TI 5510) Platforms

Codec Type	Channels per DSP in High Complexity Mode	Channels per DSP in Medium Complexity Mode	Channels per DSP in Flex Mode[1]
Low Complexity: G.711 (A-law, mu-law) Fax/modem pass-through Clear-channel	6 * 3 DSPs = 18	8 * 3 DSPs = 24	16 * 3 DSPs = 48

Table 4-3 Codec Channels Supported on the Cisco NM-HD-xx (TI 5510) Platforms (continued)

Codec Type	Channels per DSP in High Complexity Mode	Channels per DSP in Medium Complexity Mode	Channels per DSP in Flex Mode[1]
Medium Complexity: G.726 (32K, 24K, 16K) GSM Full Rate (GSMFR) Fax Relay G.729 G.729b G.729a G.729ab	6 * 3 DSPs = 18	8 * 3 DSPs = 24	8 * 3 DSPs = 24
High Complexity: G.728 G.723.1 (5.3K, 6.3K) G.723.1a (5.3K, 6.3K) GSM Enhanced Full Rate (GSMEFR) Modem Relay[2]	6 * 3 DSPs = 18	Not supported	6 * 3 DSPs = 18

1. Flex mode values are maximums for all DSPs used at the same complexity.
2. Modem relay is not currently supported in Cisco IOS Release 12.2.15ZJ.

NM-HD-*xx*

The NM-HD-*xx* can currently provide only voice termination for digital or analog interfaces. The NM-HD-*xx* uses the TI 5510 chipset, and Table 4-3 lists the number of calls that it can support in medium or high complexity mode. To determine the number of calls supported in flex mode, see the section on Flex Mode, page 4-4.

The NM-HD-1V and NM-HD-2V can be fully populated with supported voice interface cards (VICs) without oversubscribing the DSPs. The NM-HD-2VE can support any combination of analog ports or a single T1 without oversubscription. Any other combination oversubscribes the DSP resources.

Flex Mode

If the TI 5510 DSP resources are oversubscribed, you can eliminate the oversubscription by using flex mode to reallocate the DSP resources during runtime.

The NM-HD-2VE has a budget of 720 millions of instructions per second (MIPS). The DSP performs a MIPS calculation whenever a new channel becomes active, and the DSP subtracts MIPS credits from its budget whenever a new call is initiated. The number of MIPS consumed by a call depends on the type of codec used, as shown in Table 4-4. The DSP will allow a new call as long as it has more than 15 MIPS remaining in its budget before accepting the call.

Table 4-4 Flex Mode Channel Calculations for the NM-HD-xx (TI 5510) Platforms

Codec Type	MIPS Credits Required
G.711 (A-law, mu-law	15
Fax/modem pass-through	15
Clear-channel codec	15
G.726 (32K, 24K, 16K)	30

Table 4-4 Flex Mode Channel Calculations for the NM-HD-xx (TI 5510) Platforms (continued)

Codec Type	MIPS Credits Required
GSM Full Rate (GSMFR)	30
Fax Relay	30
G.729, G.729b, G.729a, G.729ab	30
G.728	40
G.723.1 (5.3K, 6.3K); G.723.1a (5.3K, 6.3K)	40
GSM Enhanced Full Rate (GSMEFR)	40
Modem Relay[1]	40

1. Modem relay is not currently supported in Cisco IOS Release 12.2.15ZJ.

Conferencing and Transcoding

This section describes how to calculate the number of DSP resources needed to support the conferencing and transcoding requirements for your IP telephony system.

NM-HDV and NM-HDV-FARM

The NM-HDV and NM-HDV-FARM provide DSP resources that can be used for voice termination, conferencing, or transcoding. The following hardware platforms can support the NM-HDV and NM-HDV-FARM modules:

- Cisco VG200
- Cisco 2600 Series
- Cisco 3620, 3640, 3645, and 3660
- Cisco 3725 and 3745

You can mix NM-HDV and NM-HDV-FARM modules in the same chassis, but not all chassis can be completely populated by these modules. Table 4-5 shows the number of modules that each type of hardware platform can support.

Table 4-5 Maximum Number of Modules per Gateway

Cisco IOS Gateway Type	Maximum NM-HDV or NM-HDV-FARM Modules per Gateway
Cisco VG200[1], 2600[2], or 3620	1
Cisco 3640	3
Cisco 3660	6
Cisco 3725[3]	2
Cisco 3745[3]	4

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610 Router. Existing models of the VG200 can still be used in an IP Telephony deployment.
2. For IP Telephony applications, use Cisco 2600XM routers. For memory considerations for the Cisco 2600 routers, see the Product Bulletin at http://www.cisco.com/warp/customer/cc/pd/rt/2600/prodlit/1675_pp.htm
3. For IP Telephony applications, the Cisco 3700 Series routers provide better performance than the Cisco 3600 Series.

Each NM-HDV or NM-HDV-FARM can hold 1 to 5 packet voice/data modules (PVDMs), and each PVDM has a fixed configuration of three DSPs.

A single DSP can support one of the following functions:

- Conferencing

  A DSP on an NM-HDV or NM-HDV-FARM can support 1 conference bridge of 1 to 6 participants.

- Transcoding

  A DSP on an NM-HDV or NM-HDV-FARM can support 4 transcoding sessions for G.711 to any other codec type except GSM. From G.711 to GSM, a single DSP can support 3 transcoding sessions.

- Voice termination

  A DSP on an NM-HDV can support 2 to 4 voice terminations, depending on the type of codec in use. (See the TI 549 column in Table 4-2.)

Calculating DSP Requirements

For sample rates of 20, 30, 40, or 60 ms with voice activity detection (VAD) enabled or disabled (or 10 ms with VAD enabled), it is possible to configure an NM-HDV or NM-HDV-FARM with a full complement of 5 PVDMs, giving 60 usable DSP resources.

To calculate the number of DSPs required for a given application, add the number of desired conferences, the number of DSPs for voice terminations, and the number of DSPs for transcoding sessions. The number of DSPs needed for transcoding is equal to the number of desired transcoding sessions divided by 4, then rounded up to the next whole number.

Because PVDMs have a fixed configuration of 3 DSPs, the number of required PVDMs is equal to the number of DSPs divided by 3, then rounded up to the next whole number.

Finally, the number of required NM-HDV or NM-HDV-FARM modules is equal to the number of PVDMs divided by 5, then rounded up to the next whole number.

For a 10 ms sampling rate with VAD disabled, it is *not* possible to utilize all DSPs on a fully populated NM-HDV. The following additional calculation is required to ensure that the packet rate does not exceed 6600 packets per second (pps), which is the capacity of the NM-HDV.

100 pps * (number of voice terminations) + 600 pps * (number of conferences) + 200 pps * (number of transcoding sessions) < 6600 pps

Conferencing Resources on Other Platforms

Table 4-6 lists the maximum number of conference participants for each codec type on the indicated platforms.

Table 4-6 Maximum Number of Conference Sessions per Platform and Codec Type

Type of Platform	Maximum Number of Participants for G.711 only	Maximum Number of Participants for G.729 only	Maximum Number of Participants for all GSM
Cisco Catalyst 4000 WS-X4604-GWY	• 24 participants total • 6 per conference bridge	• 16 participants total • 6 per conference bridge	NA
Cisco Catalyst 6000 WS-X6608-T1 or WS-X6608-E1	• 256 participants total per module • 32 per port • 6 per conference bridge	• 192 participants total per module • 24 per port • 6 per conference bridge	• 192 participants total per module • 24 per port • 6 per conference bridge
Software conference bridge (Cisco IP Voice Media Streaming Application)	• 24 participants if on the same server as Cisco CallManager • 48 participants if running on a standalone server	NA	NA

Conferencing Guidelines

This section provides conferencing resource guidelines based on the Cisco IP Telephony deployment models.

Conferencing Guidelines for All Deployment Models

- Cisco recommends that you provide the following minimum conferencing resources for your users:
  - Ad Hoc conferencing resources for at least 5% of the user base
  - Meet-Me conferencing resources for at least 5% of the user base
- In general, use media resource groups (MRGs) and media resource group lists (MRGLs) to provide sharing and load balancing of resources across multiple Cisco CallManagers. If you do not use MRGs and MRGLs, the resources are available to a single Cisco CallManager only.
- You can also use MRGs and MRGLs to separate resources based on geographical location, thereby conserving WAN bandwidth whenever possible.

Conferencing Guidelines for Single-Site Deployments

- In this model, voice traffic does not travel over an IP WAN; therefore, use a single type of codec (usually G.711), and transcoding resources are not required.
- Either software or hardware conferencing may be used. Use software conferencing for small deployments only.

Conferencing Guidelines for Multi-Site WAN Deployments with Centralized Call Processing

- In this model, call processing is localized at the central site. The MTP, transcoding, and conferencing services may be centralized or distributed, or a combination of both.
- If the media resources are centralized:
  - The WAN will be used in every call involving one of these resources.
  - Frequently, remote sites use low bit-rate (LBR) codecs across the WAN; thus, conference calls in a centralized call processing model generally require transcoding resources as well. A hardware conferencing resource is the ideal choice in this scenario because it can eliminate the need for dedicated transcoders.
  - Centrally located resources can cause local calls to traverse the WAN, so you must consider the effects on bandwidth consumption. (See Call Admission Control for Centralized Call Processing, page 1-6.)
- If the media resources are distributed:
  - Group resources into MRGLs based on their location to prevent one remote site from using resources located at another remote site. This practice helps you to manage call admission control between the sites.
  - Use hardware conferencing resources if the cluster contains more than one type of codec (for example, G.711 and G.729).

Conferencing Guidelines for Multi-Site WAN Deployments with Distributed Call Processing

- In this model, the Cisco CallManager cluster at each site may have its own conferencing resources.
- Each cluster may use multiple types of codecs, typically G.729 on the WAN between the clusters and G.711 within the cluster.
- With conferencing across multiple Cisco CallManager clusters, the identity of the conference initiator determines which conferencing resources are allocated for the conference call.
- The number of streams traversing the WAN depends on who initiates the conference and on the location of the other participants. Each participant in a cluster remote from the initiator adds an additional stream across the WAN that terminates on a resource in the initiator's cluster. Take these factors into account when configuring call admission control. (See Call Admission Control for Distributed Call Processing, page 1-12.)
- Within a single cluster, the maximum number of conference participants is 6. However, with participants from multiple clusters, more than 6 participants can be added to a conference call. Within a cluster, only the conference initiator may add a participant. However, conference participants who are members of a different cluster than the conference initiator can also add participants if they have available conference resources in their own cluster. This expansion of a conference call is possible because a cluster is not aware that an incoming call is part of an existing conference in another cluster, and each cluster can add 6 participants.

Transcoding Resources on Other Platforms

Table 4-7 lists the maximum number of media termination point (MTP) and transcoding sessions for each codec type on the indicated platforms.

Table 4-7 Maximum Number of MTP and Transcoding Sessions per Platform and Codec Type

Device Type	Codecs Supported	Maximum Number of MTP Sessions
Cisco Catalyst 4000 WS-X4604-GWY	• G.711 (A-law, mu-law) • G.729a	G.729 to G.711: • 16 MTP or transcoding sessions
Cisco Catalyst 6000 WS-X6608-T1 or WS-X6608-E1	• G.711 (A-law, mu-law) • G.723 • G.729a • GSM-FR • GSM-EFR	G.711 to any supported codec: • 24 MTP or transcoding sessions per physical port • 192 sessions per module Any single supported codec (no transcoding): • 24 MTP sessions per physical port • 192 sessions per module
Cisco ICS 7750 Multiservice Route Processor (MRP) with WAN interface cards (WICs)	• G.711 (A-law, mu-law) • G.723.1 • G.726 • G.728 • G.729a • GSM-FR • GSM-EFR	G.711 to any supported codec: • 2 MTP or transcoding sessions per digital signal processor (DSP) or packet voice/data module (PVDM) port • Maximum of 10 DSPs supported per Multiservice Route Processor (MRP) for transcoding • Total of 20 transcoding sessions per MRP
Software MTP (Cisco IP Voice Media Streaming Application)	• G.711 (A-law, mu-law)	G.711 only (no transcoding): • For 1 to 48 sessions, run the MTP software on the same server as Cisco CallManager (coresident) or on a standalone server. • For 48 to 128 sessions, run the MTP software on a standalone server.

Software MTP Resources

The following guidelines apply to software MTP resources:

- They are suited for single-site deployments, where transcoding is typically not required.
- In such deployments, the software MTP resources are needed only to support devices that are not compliant with H.323v2 (for example, Microsoft NetMeeting prior to version 3.1).
- Cisco strongly recommends running the IP Voice Media Streaming Application on a server other than the publisher or any Cisco CallManager providing call processing. The increase in CPU load from the MTP sessions might adversely impact call processing performance, and security issues can arise because User Datagram Protocol (UDP) traffic must be received on the Cisco CallManager server.

Hardware MTP and Transcoding Resources

The following guidelines apply to hardware MTP resources:

- Some resources (for example, Cisco conferencing resources) have the capability to use only G.711 voice streams.
- Use codecs other than G.711 when you want to compress the voice streams.
- When a compressed voice stream connects to a device that supports only G.711, use hardware-based MTP and transcoding services to convert the compressed voice streams into G.711.

Music on Hold

Music on hold (MoH) is an integral feature of the Cisco IP Telephony system. This feature 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 chapter describes how to design and provision MoH resources for a Cisco Enterprise IP Telephony deployment.

Cisco CallManager provides access to a variety of media resources. A *media resource* is a software-based or hardware-based entity that performs some media processing function on the voice data streams that are connected to it. Media processing functions include mixing multiple streams to create one output stream, passing the stream from one connection to another, or transcoding the data stream from one compression type to another.

Cisco CallManager allocates and uses the following types of media resources:

- Media termination point (MTP) resources
- Transcoding resources
- Unicast Conferencing Resources
- Music on hold resources

For more information about media resources in general, see the chapter on Media Resources, page 4-1.

This chapter examines the following design aspects of the MoH feature:

Deployment Basics of MoH

For callers to hear music on hold, Cisco CallManager must be configured to support the MoH feature. The MoH feature has two main requirements:

- An MoH server to provide the MoH audio stream sources
- 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 possible 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 interactive voice response (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 two types of MoH transport mechanisms:

- Unicast
- Multicast

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 music on hold 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. Thus, 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 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. 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 music on hold conserves system resources and bandwidth because it enables multiple users to use the same audio source stream to provide music on hold. Thus, 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, refer to the *Cisco AVVID Network Infrastructure IP Multicast Design* document, available online at

http://cisco.com/go/srnd

Note The following 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 2600, 3600, and 3700 Series Routers with MGCP or H.323 and Cisco IOS Release 12.2(8)T or later.

Coresident and Standalone MoH Servers

The MoH feature requires the use of a server that is part of a Cisco CallManager cluster. You can configure the MoH server in either of the following ways:

- Coresident deployment

  In a coresident deployment, the MoH feature runs on any server (either publisher or subscriber) in the cluster that is also running the Cisco CallManager software. Because MoH shares server resources with Cisco CallManager in a coresident configuration, this type of configuration drastically reduces the number of simultaneous streams that an MoH server can send.

- Standalone deployment

  A standalone deployment places the MoH feature on a dedicated server within the Cisco CallManager cluster. The sole function of this dedicated server is to send MoH streams to devices within the network. A standalone deployment allows for the maximum number of streams from a single MoH server.

Fixed and Audio File MoH Sources

You can set up the source for MoH in any of the following ways:

- MoH from an audio file on the Cisco CallManager or MoH server
  - Unicast MoH from an audio file
  - Multicast MoH from an audio file
- MoH from a fixed music source (via sound card)
  - Unicast MoH from a fixed source
  - Multicast MoH from a fixed source

MoH can be generated from an audio file stored on the MoH server. Audio files must be in one of the following formats:

- G.711 A-law or mu-law (recorded at a sampling rate of 8 KHz)
- G.729 Annex A
- Wideband

You can create these files with the Cisco MoH Audio Translator service, which transcodes and formats audio source files (such as .wav or .mp3 files) into the appropriate MoH source file for the specified codec type(s). The MoH server requests these files based on the audio sources configured and loads them into memory during initialization or when the audio sources are requested. When an MoH event occurs, the configured audio source file is streamed to the requesting device on hold.

If recorded or live audio is needed, MoH can be generated from a fixed source. For this type of MoH, a sound card is required. The fixed audio source is provided by the Microsoft Windows audio input that is normally linked to the local sound card.

This mechanism enables you to use radios, CD players, or any other compatible sound source. The stream from the fixed audio source is transcoded in real-time to support the codec that was configured through Cisco CallManager Administration. The fixed audio source can be transcoded into G.711 (A-law or mu-law), G.729 Annex A, and Wideband, and it is the only audio source that is transcoded in real-time.

The following sound cards are supported for fixed or live audio source:

- Soundblaster PCI 16
- Telex P-800 USB

  A USB sound device supported in Cisco CallManager Release 3.3(3) with Windows 2000 (OS 2000 version 2.5). Only the following platforms support this device: MCS-7825H-2.2-EVV1, MCS-7835H-2.4-EVV1, MCS-7835I-2.4-EVV1 MCS-7845H-2.4-EVV1, and MCS-7835-1266.

Note Prior to using a fixed audio source to transmit music on hold, you should consider the legalities and the ramifications of re-broadcasting audio materials. Consult your legal department for potential issues.

MoH Server as Part of the Cisco CallManager Cluster

The MoH feature requires that each MoH server must be part of a Cisco CallManager cluster. All MoH servers must share their configurations with the publisher server and participate in the SQL replication schema. Specifically, the MoH server must share the following information (configured through Cisco CallManager Administration) by means of the SQL database:

- Audio sources — The number and identity of all configured MoH audio sources
- Multicast or unicast — The transport nature (multicast or unicast) configured for each of these sources
- Multicast address — The multicast base IP address of those sources configured to stream as multicast

The MoH server becomes part of the Cisco CallManager cluster and participates in the SQL database replication automatically. To configure a standalone MoH server, start with a normal Cisco CallManager installation on that server. Then disable the Cisco CallManager service (on the standalone MoH server only) and enable the Cisco IP Voice Media Streaming Application.

Basic MoH and MoH Call Flows

This section describes the basic operation of MoH as implemented in Cisco CallManager as well as typical call flow scenarios.

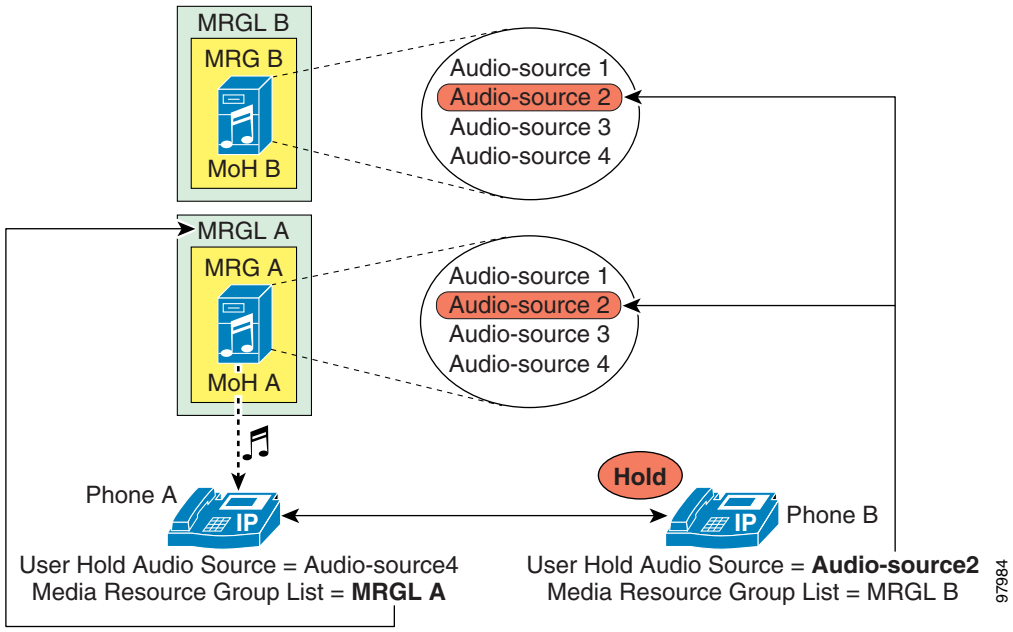
Basic MoH

The basic operation of MoH in a Cisco IP Telephony environment consists of a holder and holdee. The *holder* is the endpoint user or network application placing a call on hold, and the *holdee* is the endpoint user or device placed on hold.

The MoH stream that an endpoint receives is determined by a combination of the User Hold MoH Audio Source of the device placing the endpoint on hold (holder) and the configured media resource group list (MRGL) of the endpoint placed on hold (holdee). The User Hold MoH Audio Source configured for the holder determines the audio file that will be streamed when the holder puts a call on hold, and the holdee's configured MRGL indicates the resource or server from which the holdee will receive the MoH stream.

In simplest terms, the holder's configuration determines which audio file to play, and the holdee's configuration determines which resource or server will play that file. As illustrated by the example in Figure 5-1, if phones A and B are on a call and phone B (holder) places phone A (holdee) on hold, phone A will hear the MoH audio source configured for phone B (Audio-source2). However, phone A will receive this MoH audio stream from the MRGL (resource or server) configured for phone A (MRGL A).

Figure 5-1 User Hold Audio Source and Media Resource Group List (MRGL)



Because the configured MRGL determines the server from which a unicast-only device will receive the MoH stream, you must configure unicast-only devices with an MRGL that points to a unicast MoH resource or media resource group (MRG). Likewise, a device capable of multicast should be configured with an MRGL that points to a multicast MRG.

MoH Configuration Settings

You can configure the settings for MRGLs and User and Network Hold Audio Sources in several places within Cisco CallManager Administration, and you can configure different (and potentially conflicting) settings in each place.

To determine which User and Network Audio Source configuration setting to apply in a particular case, Cisco CallManager interprets these settings for the *holder* device in the following priority order:

1. Directory or line setting (Devices with no line definition, such as gateways, do not have this level.)
2. Device setting
3. Device pool setting
4. Cluster-wide default setting

When attempting to determine the audio source for a particular holder, Cisco CallManager first looks at the User Audio Source configured at the directory or line level. If this level is not defined, then Cisco CallManager looks at the User Audio Source configured on the holder device. If this level is not defined, then Cisco CallManager looks at the User Audio Source configured for the device pool of the

holder device. If this level is not defined, then Cisco CallManager looks at the cluster-wide default audio source ID configured under the Cisco CallManager system parameters. (By default, this audio source ID is set to 1, which is the SampleAudioSource.)

Cisco CallManager also interprets the MRGL configuration settings of the *holdee* device in the following priority order:

1. Device setting
2. Device pool setting
3. System default MoH resources

When attempting to determine the MRGL for a particular holdee, Cisco CallManager looks at the MRGL configured at the device level. If this level is not defined, then Cisco CallManager looks at the MRGL configured for the device pool of the holdee device. If this level is not defined, then Cisco CallManager uses the system default MoH resources. System default MoH resources are those resources not assigned to any MRG, and they are always unicast.

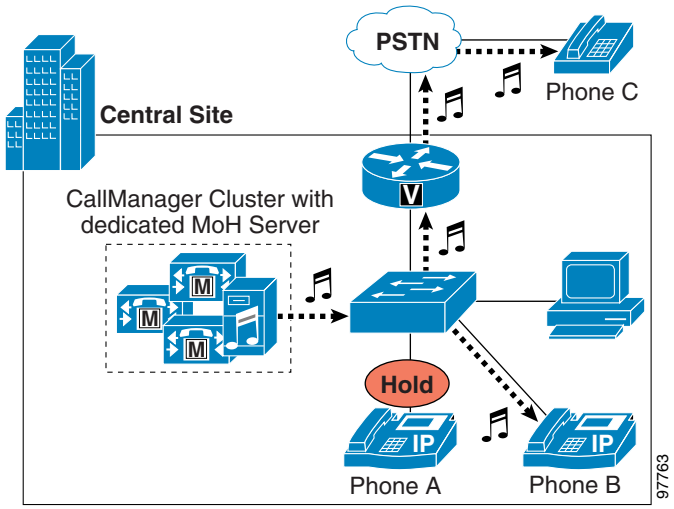
User and Network Hold

There are two basic types of user hold:

- User on hold at an IP phone or other endpoint device
- User on hold at the PSTN, where MoH is streamed to the gateway

Figure 5-2 shows these two types of call flows. If phone A is in a call with phone B and phone A (holder) pushes the Hold softkey, then a music stream is sent from the MoH server to phone B (holdee). The music stream can be sent to holdees within the IP network or holdees on the PSTN, as is the case if phone A places phone C on hold. In the case of phone C, the MoH stream is sent to the voice gateway interface and converted to the appropriate format for the PSTN phone. When phone A presses the Resume softkey, the holdee (phone B or C) disconnects from the music stream and reconnects to phone A.

Figure 5-2 Basic User Hold Example

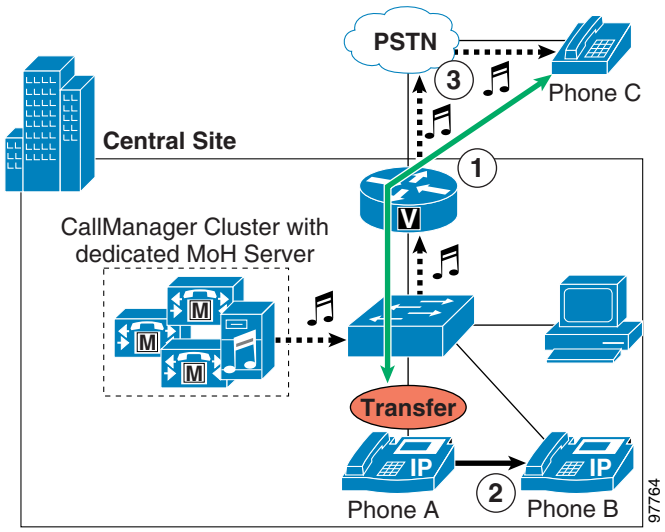


Network hold includes the following types:

- Call transfer
- Call Park
- Conference setup
- Application-based hold

Figure 5-3 shows the call transfer call flow. When phone A receives a call from the PSTN phone C (step 1), phone A answers the call and then transfers it to phone B (step 2). During the transfer process, phone C receives an MoH stream from the MoH server via the gateway (step 3). After phone A completes the transfer action, phone C disconnects from the music stream and gets redirected to phone B, the transfer destination. This process is the same for other network hold operations such as call park and conference setup.

Figure 5-3 Basic Network Hold Example for Call Transfer



Unicast and Multicast MoH Call Flows

MoH operation is quite similar to normal phone call flows, with the MoH server acting as a Skinny Client Control Protocol (SCCP) device to which the holdee device can connect or disconnect as required. However, there are distinct differences between unicast and multicast MoH call flow behavior. A unicast MoH call flow is initiated by a message from Cisco CallManager to the MoH server. This message tells the MoH server to send an audio stream to the holdee device's IP address. On the other hand, a multicast MoH call flow is initiated by a message from Cisco CallManager to the holdee device. This message instructs the endpoint device to join the multicast group address of the configured multicast MoH audio stream.

For a detailed look at MoH call flows, see the section on Detailed Unicast and Multicast MoH Call Flows, page 5-17.

MoH Configuration Considerations and Best Practices

This section highlights some MoH configuration considerations and best practice to help you design a robust MoH solution.

Codec Selection

If you need multiple codecs for MoH deployment, configure them in the IP Voice Streaming Media App service parameter under Cisco CallManager Service Parameters Configuration. Select the desired codec types from the Supported MoH Codecs list under the Clusterwide Parameters section. By default, only G.711 mu-law is selected. To select another codec type, click on it in the scrollable list. For multiple selections, hold down the CTRL key and use the mouse to select multiple codecs from the scrollable list. After making your selection, click the Update button.

Note If you are using the G.729 codec for MoH audio streams, be aware that this codec is optimized for speech and it provides only marginal audio fidelity for music.

Multicast Addressing

Proper IP addressing is important for configuring multicast MoH. Addresses for IP multicast range from 224.0.1.0 to 239.255.255.255. The Internet Assigned Numbers Authority (IANA), however, assigns addresses in the range 224.0.1.0 to 238.255.255.255 for public multicast applications. Cisco strongly discourages using public multicast addresses for music on hold. Instead, Cisco recommends that you configure multicast MoH audio sources to use IP addresses in the range 239.1.1.1 to 239.255.255.255, which is reserved for administratively controlled applications on private networks.

Furthermore, you should configure multicast audio sources to increment on the IP address and not the port number, for the following reasons:

- IP phones placed on hold join multicast IP addresses, not port numbers.

  Cisco IP phones have no concept of multicast port numbers. Therefore, if all the configured codecs for a particular audio stream transmit to the same multicast IP address (even on different port numbers), all streams will be sent to the IP phone even though only one stream is needed. This has the potential of saturating the network with unnecessary traffic because the IP phone is capable of receiving only a single MoH stream.

- IP network routers route multicast based on IP addresses, not port numbers.

  Routers have no concept of multicast port numbers. Thus, when it encounters multiple streams sent to the same multicast group address (even on different port numbers), the router forwards all streams of the multicast group. Because only one stream is needed, network bandwidth is over-utilized and network congestion can eventually result.

MoH Audio Sources

Audio sources are shared among *all* MoH servers in the Cisco CallManager cluster. You can configure up to 51 unique audio sources per cluster (50 audio file sources and one fixed/live source). For exceptions to this limit, refer to the sections on Using Multiple Fixed or Live Audio Sources, page 5-9 and Multicast MoH from Branch Router Flash, page 5-14.

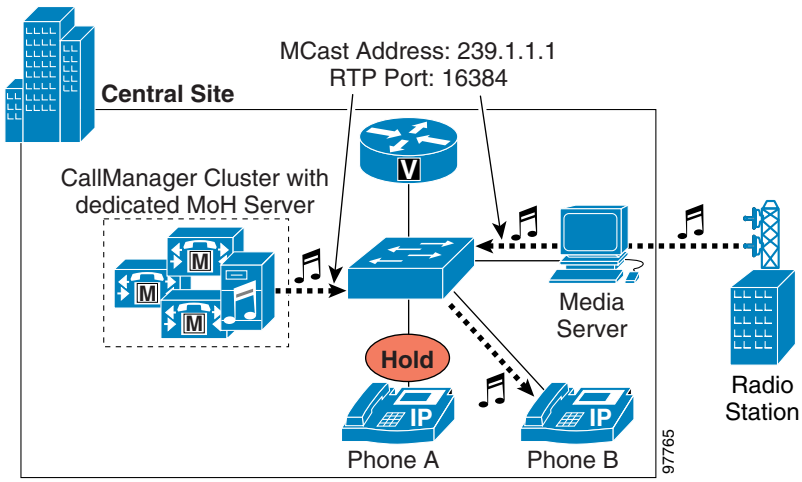
Using Multiple Fixed or Live Audio Sources

Each MoH server is capable of streaming only one fixed audio source. In most cases, if multiple fixed or live audio sources are needed, a separate MoH server is required for each source. However, it is possible to provide multiple fixed-source MoH audio streams by using external non-MoH servers or devices that are capable of streaming multicast from fixed or live sources.

For each external source, you must configure the MoH server with an audio source that has the same multicast IP address and port number as that of the audio source stream being multicast by the external source server or device. In addition, you should block this configured (non-external) audio source from traversing the WAN by setting its maximum hop count to one (1) or by using access control lists (ACLs) to disallow the packets from streaming further than the local subnet.

Figure 5-4 shows an example of an external live source being used as an MoH stream. In this figure, the MoH server is streaming a multicast audio source to 239.1.1.1 (on RTP port 16384), and this stream has been limited to a maximum hop of one, thus ensuring that it will not travel off the local MoH server's subnet. At the same time, the media server is multicasting an audio stream derived from a live radio station feed. This stream is also using 239.1.1.1 as its multicast address and 16384 as the RTP port number, but this stream has a hop-count or TTL greater than one to ensure that it is able to reach phone B when phone A presses the Hold softkey.

Figure 5-4 External Live Audio Source Example



Note Using live radio broadcasts as multicast audio sources can have legal ramifications. Consult your legal department for potential issues.

Numerous streams can be multicast from one or more external media servers, by configuring additional audio sources on multiple MoH servers and then sourcing audio streams from the external servers using the same multicast group addresses configured on the MoH servers. However, because a combination of the user/network hold audio source of the holder and the MRGL of the holdee determines the MoH stream that an endpoint device hears, it can become difficult to predict which particular stream an endpoint will receive in an environment with many overlapping multicast group addresses. For this reason, Cisco recommends that you configure only a single multicast audio source on each MoH server. This recommendation ensures that the audio source an endpoint receives is uniquely identifiable by a single combination of user/network hold audio source and MRGL.

Unicast and Multicast in the Same Cisco CallManager Cluster

In some cases, administrators might want to configure a single Cisco CallManager cluster to handle both unicast and multicast MoH streams. This configuration might be necessary because the telephony network contains devices or endpoint that do not support multicast or because some portions of the network are not enabled for multicast.

Use one of the following methods to enable a cluster to support both unicast and multicast MoH audio streams:

- Deploy separate MoH servers, with one server configured as a unicast MoH server and the second server configured as a multicast MoH server.
- Configure separate media resource groups (MRGs) for the same MoH server, with one MRG configured to use multicast for audio streams and the second MRG configured to use unicast.

In either case, you must configure at least two MRGs and at least two media resource group lists (MRGLs). Configure one unicast MRG and one unicast MRGL for those endpoints requiring unicast MoH. Likewise, configure one multicast MRG and one multicast MRGL for those endpoints requiring multicast MoH.

When deploying separate MoH servers, configure one server without multicast enabled (unicast-only) and configure a second MoH server with multicast enabled. Assign the unicast audio resource of the unicast-only MoH server and the multicast audio resource of the multicast MoH server to the unicast and multicast MRGs, respectively. Ensure that the **Use Multicast for MoH Audio** box is checked for the multicast MRG but not for the unicast MRG. Also assign these unicast and multicast MRGs to their respective MRGLs. In this case, an MoH stream is unicast or multicast based on the server from which it is served and on whether the MRG is configured to use multicast.

When configuring separate MRGs for the same MoH server, configure the server and its audio source for multicast. Assign this same audio source to both the unicast MRG and the multicast MRG, and check the **Use Multicast for MoH Audio** box for the multicast MRG. In this case, an MoH stream is unicast or multicast based solely on whether the MRG is configured to use multicast.

Note Configuring the unicast MRG can be confusing because the audio resource you are adding to this MRG has [Multicast] appended to the end of the resource name even though you are adding it to the unicast MRG. This label is simply an indication that the resource is capable of being multicast, but the Use Multicast for MoH Audio box determines whether the resource will be sent as unicast or multicast.

In addition, you must configure individual devices or device pools to use the appropriate MRGL. You can place all unicast devices in a device pool or pools and configure those device pools to use the unicast MRGL. Likewise, you can place all multicast devices in a device pool or pools and configure those device pools to use the multicast MRGL. Optionally, you can configure individual devices to use the appropriate unicast or multicast MRGL. Also configure a User Hold Audio Source and Network Hold Audio Source for each device pool, individual device, or (in the case of phone devices) individual lines or directory numbers to determine the appropriate audio source to stream.

Redundancy

Cisco recommends that you configure and deploy multiple MoH servers for completely redundant MoH operation. If the first MoH server fails or becomes unavailable because it no longer has the resources required to service requests, the second server can provide continued MoH functionality. For proper redundant configuration, assign resources from at least two MoH servers to each MRG in the cluster.

Resources in the MRG are used in the order listed. When a device requests an MoH audio resource, Cisco CallManager attempts to stream the first MoH resource in the MRG to the device. If the first resource is unavailable (due to server failure or lack of resources), Cisco CallManager then attempts to use the next MoH resource in the MRG.

In environments where both multicast and unicast MoH are required, be sure to provide redundancy for both transport types to ensure MoH redundancy for all endpoints in the network.

Quality of Service (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 will receive the same classification and priority queueing treatment as voice RTP media traffic.

Hardware and Capacity Planning for MoH Resources

As with all media resources, capacity planning is crucial to make certain that the hardware, once deployed and configured, can support the anticipated call volume of the network. For this reason, it is important to be aware of the hardware capacity for MoH resources and to consider the implications of multicast and unicast MoH in relation to this capacity.

Server Platform Limits

Table 5-1 lists the server platforms and the maximum number of simultaneous MoH sessions each can support. Ensure that network call volumes do not exceed these limits because, once MoH sessions have reached these limits, additional load could result in poor MoH quality, erratic MoH operation, or even loss of MoH functionality.

Table 5-1 Maximum Number of MoH Sessions per Server Platform Type

Server Platform	Codecs Supported	MoH Sessions Supported
MCS 7815 MCS 782x (All models) MCS 7830 (All models) SPE-310 HP DL320 IBM xSeries 33x (All models)	G.711 (A-law and mu-law) G.729a Wideband audio	Coresident server: 20 MoH sessions Standalone MoH server: 50 MoH sessions
MCS 7835 (All models) MCS 7845 (All models) HP DL380 IBM xSeries 34x (All models)	G.711 (A-law and mu-law) G.729a Wideband audio	Coresident server: 20 MoH sessions Standalone MoH server: 250 MoH sessions[1]

1. You can configure a maximum of 51 unique audio sources per Cisco CallManager cluster.

Note The maximum session limits listed in Table 5-1 apply to unicast, multicast, or simultaneous unicast and multicast sessions. The limits represent the recommended maximum sessions a platform can support, irrespective of the transport mechanism.

Resource Provisioning and Capacity Planning

When provisioning for coresident or standalone MoH server configurations, network administrators should consider the type of transport mechanism used for the MoH audio streams. If using unicast MoH, each device on hold will require a separate MoH stream. However, if using multicast MoH and only a single audio source, then only a single MoH stream will be required no matter how many devices are on hold.

For example, given a cluster with 30,000 phones and a 2% hold rate (only 2% of all endpoint devices are on hold at any given time), 600 MoH streams or sessions would be required. Given a unicast-only MoH environment, two standalone MoH servers (MCS 7835 or 7845) and five coresident MoH servers would be required to handle this load, as shown by the following calculation:

[(250 sessions per MCS 7835 or 7845 standalone server) * (2 standalone servers)] + [(20 sessions per coresident server) * (5 coresident servers)] = 600 sessions

By comparison, a multicast-only MoH environment with 40 unique MoH audio streams, for example, would require only two coresident MoH servers, as shown by the following calculation:

(20 sessions per coresident server) * (2 coresident servers) = 40 sessions

As these examples show, multicast MoH can provide a considerable savings in server resources over unicast MoH.

In the preceding examples, the 2% hold rate is based on 30,000 phones and does not take into account gateways or other endpoint devices in the network that are also capable of being placed on hold. You should consider these other devices when calculating a hold rate because they could potentially be placed on hold just as the phones can.

The preceding calculations also do not provide for MoH server redundancy. If an MoH server fails or if more than 2% of the users go on hold at the same time, there are no other MoH resources in this scenario to handle the overflow or additional load. Your MoH resource calculations should include enough extra capacity to provide for redundancy.

Implications for MoH With Regard to IP Telephony Deployment Models

The various IP Telephony deployment models introduce additional considerations for MoH configuration design. Which deployment model you choose can also affect your decisions about MoH transport mechanisms (unicast or multicast), resource provisioning, and codecs. This section discusses these issues in relation to the various deployment models.

For more detailed information about the deployment models, see the chapter on IP Telephony Deployment Models, page 1-1.

Single-Site Campus (Relevant to All Deployments)

Single-site campus deployments are typically based on a LAN infrastructure and provide sufficient bandwidth for large amounts of traffic. Because bandwidth is typically not limited in a LAN infrastructure, Cisco recommends the use of the G.711 (A-law or mu-law) codec for all MoH audio streams in a single-site deployment. G.711 provides the most optimal voice and music streaming quality in an IP Telephony environment.

MoH server redundancy should also be considered. In the event that an MoH server becomes overloaded or is unavailable, configuring multiple MoH servers and assigning them in preferred order to MRGs ensures that another server can take over and provide the MoH streams.

With the increasing diversity of network technologies, in a large single-site campus it is likely that some endpoint devices will be unable to support multicast. For this reason, you might have to deploy both unicast and multicast MoH resources. For example, wireless IP phones do not support multicast due to the behavior of wireless technology. Thus, when deploying wireless IP phones, you have to configure both multicast and unicast MoH.

To ensure that off-net calls and application-handled calls receive expected MoH streams when placed on hold, configure all gateways and other devices with the appropriate MRGLs and audio sources or assign them to appropriate device pools.

Centralized Multi-Site Deployments

Multi-site IP telephony deployments with centralized call processing typically contain WAN connections to multiple non-central sites. These WAN links usually cause bandwidth and throughput bottlenecks. To minimize bandwidth consumption on these links, Cisco recommends the use of the G.729 codec for all MoH audio streams traversing the WAN. Because the G.729 codec is optimized for voice and not music applications, you should use G.729 only across the WAN, where the bandwidth savings far outweighs the lower quality afforded by G.729 for MoH transport. Likewise, because multicast traffic provides significant bandwidth savings, you should always use multicast MoH when streaming audio to endpoints across the WAN.

Call Admission Control and MoH

Call admission control (CAC) is required when IP telephony traffic is traveling across WAN links. Due to the limited bandwidth available on these links, it is highly possible that voice media traffic might get delayed or dropped. The Cisco CallManager locations-based call admission control mechanism enables you to configure each location in the IP telephony environment to accept or allow only a certain number of calls across the WAN link to other locations, thus preventing over-subscription of WAN bandwidth and delayed or lost voice packets. By specifying a bandwidth value for the WAN link, you can limit the number of calls based on the speed of the link. If the limit is reached or exceeded, Cisco CallManager rejects all other calls that are attempted across the link.

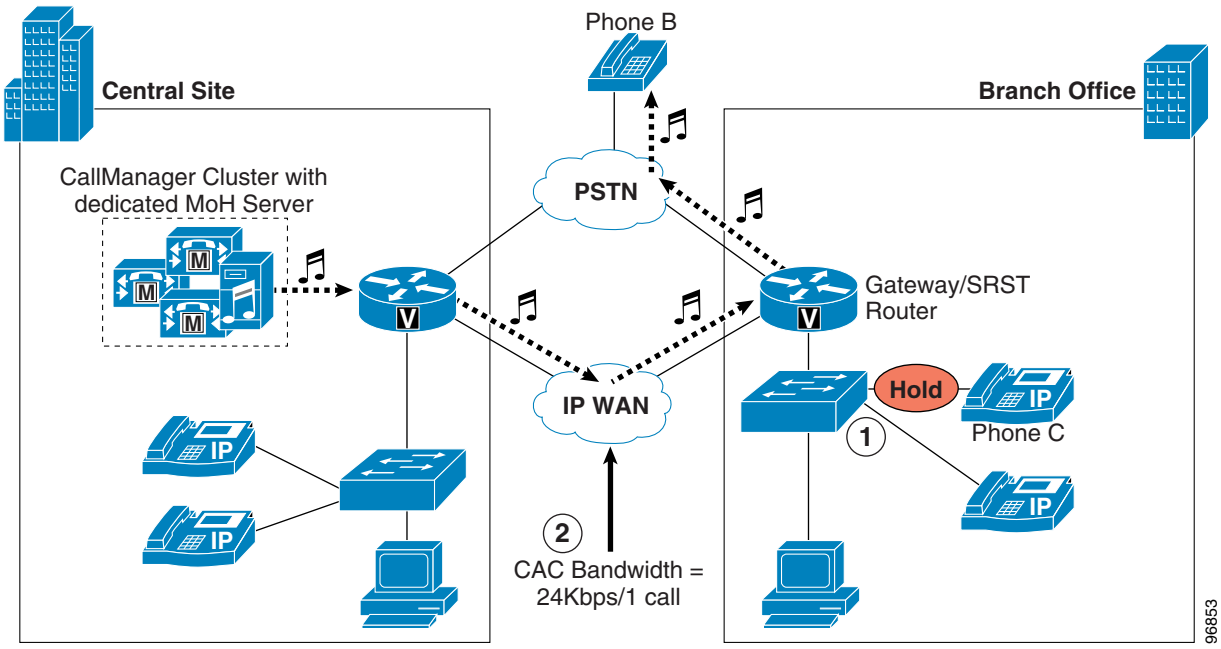
Cisco CallManager locations-based call admission control is capable of tracking unicast MoH streams traversing the WAN but not multicast MoH streams. Thus, even if WAN bandwidth has been fully subscribed, a multicast MoH stream will not be denied access to the WAN by call admission control. Instead, the stream will be sent across the WAN, likely resulting in poor audio stream quality and poor quality on all other calls traversing the WAN. To ensure that multicast MoH streams do not cause this over-subscription situation, you should over-provision the QoS configuration on all WAN interfaces by configuring the low-latency queuing (LLQ) voice priority queue with additional bandwidth. Add enough

bandwidth for every unique multicast MoH stream that might traverse the WAN link. For example, if there are four unique multicast audio streams that could potentially traverse the WAN, then add 96 Kbps to the voice priority queue (4 * 24 Kbps (for G.729 audio stream) = 96 Kbps).

Figure 5-5 shows an example of call admission control and MoH in a centralized multi-site deployment. For this example, assume that phone C is in a call with a PSTN phone (phone B). At this point, no bandwidth has been consumed on the WAN. When phone C pushes the Hold softkey (step 1), phone B receives an MoH stream from the central-site MoH server via the WAN, thereby consuming bandwidth on the link. Whether or not this bandwidth is taken into consideration by call admission control depends on the type of MoH stream. If multicast MoH is streamed, then call admission control will not consider the 24 Kbps being consumed (therefore, QoS on the WAN interfaces should be provisioned accordingly). However, if unicast MoH is streamed, call admission control will subtract 24 Kbps from the available WAN bandwidth (step 2).

Note The preceding example might seem to imply that unicast MoH should be streamed across the WAN. This is merely an example used to illustrate locations-based call admission control with MoH and is not intended as a recommendation or endorsement of this configuration. As stated earlier, multicast MoH is the recommended transport mechanism for sending MoH audio streams across the WAN.

Figure 5-5 Locations-Based Call Admission Control and MoH



Multicast MoH from Branch Router Flash

Beginning with Cisco IOS Release 12.2(15)ZJ and SRST Release 3.0, MoH can be multicast in a remote or branch site via the branch router's flash. Multicast MoH from a Cisco IOS router's flash enhances the MoH feature for the following reasons:

- The branch gateway or router can provide multicast MoH when it is in SRST mode and the branch devices have lost connectivity to the central-site Cisco CallManager.

- This configuration can eliminate the need to forward MoH across the WAN to remote branch sites by providing locally sourced MoH even when the WAN is up and the phones are controlled by Cisco CallManager.

Example 5-1 illustrates the commands to use in the Cisco IOS router configuration (under the SRST section) to enable multicast MoH from the router flash:

Example 5-1 Enabling Multicast MoH from Branch Router Flash

```
SRST-router(config)#call-manager-fallback
SRST-router(config-cm-fallback)#moh music-on-hold.au
SRST-router(config-cm-fallback)#multicast moh 239.192.240.1 port 16384 route 10.1.1.254
```

In Example 5-1, the name of the audio file on the router flash is music-on-hold.au, and the configured multicast address and port number are 239.192.240.1 and 16384 respectively. The optional **route** command indicates a source interface address for the multicast stream. If no route option is specified, the multicast stream will be sourced from the configured SRST default address. Note that you can stream only a single audio file from flash and that you can use only a single multicast address and port number per router.

When the branch router is operating in SRST mode, it can stream multicast MoH to all analog and digital ports within the chassis, thereby providing MoH to analog phones and PSTN callers. At this time, IP phones in SRST mode cannot receive multicast MoH from the SRST router's flash and will receive tone-on-hold instead.

Once configured, the router will continue to stream the MoH stream from flash even when not in SRST mode. When the branch router is not operating in SRST mode, it can multicast MoH from the flash to all local devices, including IP phones. The branch router's configuration for non-SRST multicast MoH from flash is the same as for the SRST configuration. (See Example 5-1.) However, which multicast address you configure on the router depends on the intended operation. If you want multicast MoH from flash only in SRST mode (for example, if MoH received by remote devices is sourced from the central MoH server when not in SRST mode), then the multicast address and port number configured on the router should not overlap with any of the central-site MoH server audio sources. Otherwise, remote devices might continue to receive MoH from the local router flash, depending on the configured user/network hold audio sources.

If you always want multicast MoH from the branch router flash, then you must configure the central-site server with an audio source that has the same multicast IP address and port number as configured on the branch router. In this scenario, because the multicast MoH audio stream is always coming from the router's flash, it is not necessary for the central site MoH server audio source to traverse the WAN.

To prevent the central site audio stream(s) from traversing the WAN, use either of the following methods:

- Maximum hop count configuration

  Configure the central-site MoH audio source with a maximum hop count (or TTL) low enough to ensure that it will not stream further than the central-site LAN.

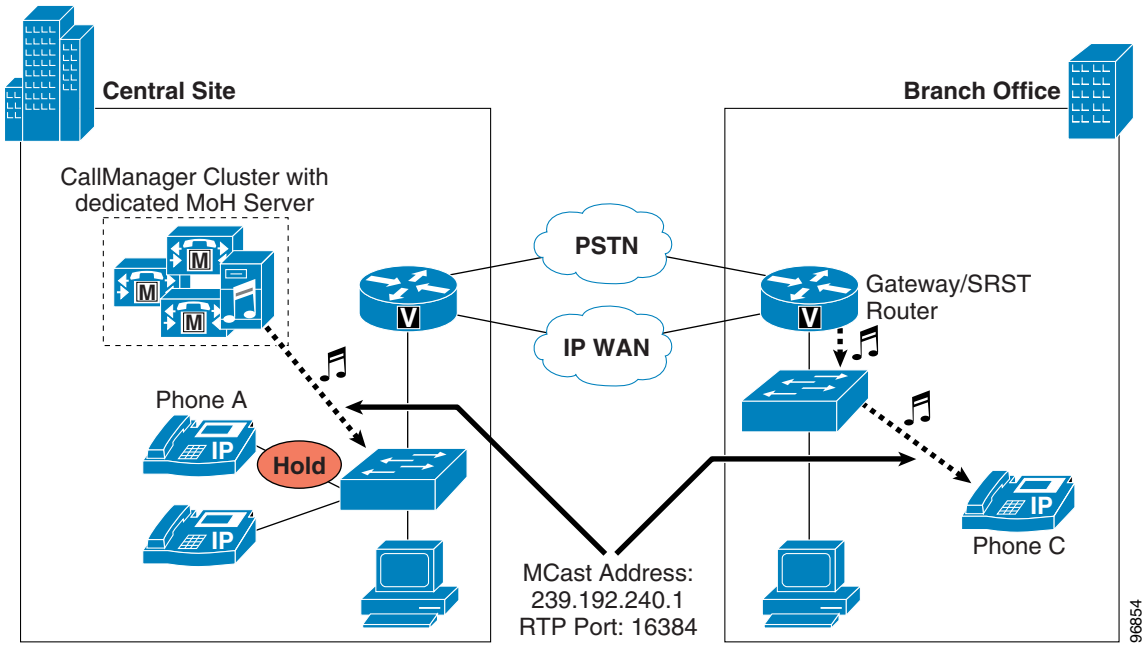
- Access control list (ACL) on the WAN interface

  Configure an ACL on the central-site WAN interface to disallow packets destined to the multicast group address(es) from being sent out the interface.

Figure 5-6 illustrates streaming multicast MoH from the flash of a remote router when it is not in SRST mode. After phone A places phone C on hold, phone C receives multicast MoH from the local SRST router. In this figure, the MoH server is streaming a multicast audio source to 239.192.240.1 (on RTP port 16384), however this stream has been limited to a maximum hop of one (1) to ensure that it will not travel off the local MoH server's subnet and across the WAN. At the same time, the branch office SRST

router/gateway is multicasting an audio stream from flash. This stream is also using 239.192.240.1 as its multicast address and 16384 as the RTP port number. When phone A presses the Hold softkey, phone C receives the MoH audio stream sourced by the SRST router.

Figure 5-6 Multicast MoH from Branch Router Flash



When using this method for delivering multicast MoH, configure all devices within the Cisco CallManager cluster to use the same user hold and network hold audio source and configure all branch routers with the same multicast group address and port number. Because the user or network hold audio source of the holder is used to determine the audio source, if you configure more than one user or network hold audio source within the cluster, there is no way to guarantee that a remote holdee will always receive the local MoH stream. For example, suppose a central-site phone is configured with an audio source that uses group address 239.192.254.1 as its user and network hold audio source. If this phone places a remote device on hold, the remote device will attempt to join 239.192.254.1 even if the local router flash MoH stream is sending to multicast group address 239.192.240.1. If instead all devices in the network are configured to use the user/network hold audio source with multicast group address 239.192.240.1 and all branch routers are configured to multicast from flash on 239.192.240.1, then every remote device will receive the MoH from its local router's flash.

In networks with multiple branch routers configured to stream multicast MoH from flash, it is possible to have more than 51 unique MoH audio sources in a cluster. Each branch site router can multicast a unique audio file from flash, although all routers must multicast this audio on the same multicast group address. In addition, the central-site MoH server can multicast a MoH stream on this same multicast group address. Thus, if there are 100 branch sites each multicasting an audio file from flash, then the cluster can contain 101 unique MoH audio sources (100 branch streams and one central-site stream). If you want more than one unique audio stream in the central site, you can stream fixed/live sources from additional MoH servers or from external media servers (as described in Using Multiple Fixed or Live Audio Sources, page 5-9), but you should not configure more than one audio source per server.

Distributed Multi-Site Deployments

Multi-site IP telephony deployments with distributed call processing typically contain WAN or MAN connections between the sites. These lower-speed links usually cause bandwidth and throughput bottlenecks. To minimize bandwidth consumption on these links, Cisco recommends the use of G.729 codec for all MoH audio streams traversing them. Because the G.729 codec is optimized for voice and not music applications, you should use G.729 only across the WAN/MAN links, where the bandwidth savings far outweighs the lower quality afforded by G.729 for MoH transport.

Multicast MoH is not supported for calls between Cisco CallManager clusters (intercluster calls). Therefore, you must configure at least one unicast MoH resource in each Cisco CallManager cluster if you want MoH on the intercluster trunk (ICT).

Proper multicast address management is another important design consideration in the distributed intercluster environment. All MoH audio source multicast addresses must be unique across all Cisco CallManager clusters in the deployment to prevent possible overlap of streaming resources throughout the distributed network.

Clustering Over the WAN

As its name suggests, clustering-over-the-WAN deployments also contain the same type of lower-speed WAN links as other multi-site deployments and therefore are subject to the same requirements for G.729 codec, multicast transport mechanism, and solid QoS for MoH traffic traversing these links is imperative.

In addition, you should deploy MoH server resources at each side of the WAN in this type of configuration. In the event of a WAN failure, devices on each side of the WAN will be able to continue to receive MoH audio streams from their locally deployed MoH server. Furthermore, proper MoH redundancy configuration is extremely important. The devices on each side of the WAN should point to an MRGL whose MRG has a priority list of MoH resources with at least one local resource as the highest priority. Additional MoH resources should be configured for this MRG in the event that the primary server becomes unavailable or is unable to process requests. At least one other MoH resource in the list should point to an MoH resource on the remote side of the WAN in the event that resources at the local side of the WAN are unavailable.

Detailed Unicast and Multicast MoH Call Flows

Figure 5-7 illustrates a typical multicast call flow. As shown in the diagram, when the Hold softkey is pressed at phone A, Cisco CallManager instructs both phone A and phone B to Close Receive Channel and Stop Media Transmission. This action effectively stops the RTP two-way audio stream. Next, Cisco CallManager tells phone B (the holdee) to Start Multicast Media Reception from multicast group address 239.192.240.1. The phone then issues an Internet Group Management Protocol (IGMP) membership report indicating that it is joining this group.

Meanwhile, the MoH server has been sourcing RTP audio to this multicast group address and, upon joining the multicast group, phone B begins receiving the MoH stream. Once phone A presses the Resume softkey, Cisco CallManager instructs phone B to Stop Multicast Media Reception, effectively ending the MoH session. Next, Cisco CallManager sends a series of Open Receive Channel messages to phones A and B, just as would be sent at the beginning of a phone call between the two phones. Soon afterwards, Cisco CallManager instructs both phones to Start Media Transmission to each other's IP addresses (in this case, 10.96.200.12 and 10.96.200.13). The phones are once again connected via an RTP two-way audio stream.

Note The call flow diagrams in Figure 5-7 and Figure 5-8 assume that an initial call is up between phones A and B, with a two-way RTP audio stream. These diagrams are representative of call flows and therefore include only the pertinent traffic required for proper MoH operation. Thus, keepalives, acknowledgements, and other miscellaneous traffic have been eliminated to better illustrate the interaction. The initial event in each diagram is the Hold softkey action performed by phone A.

Figure 5-7 Detailed Multicast MoH Call Flow

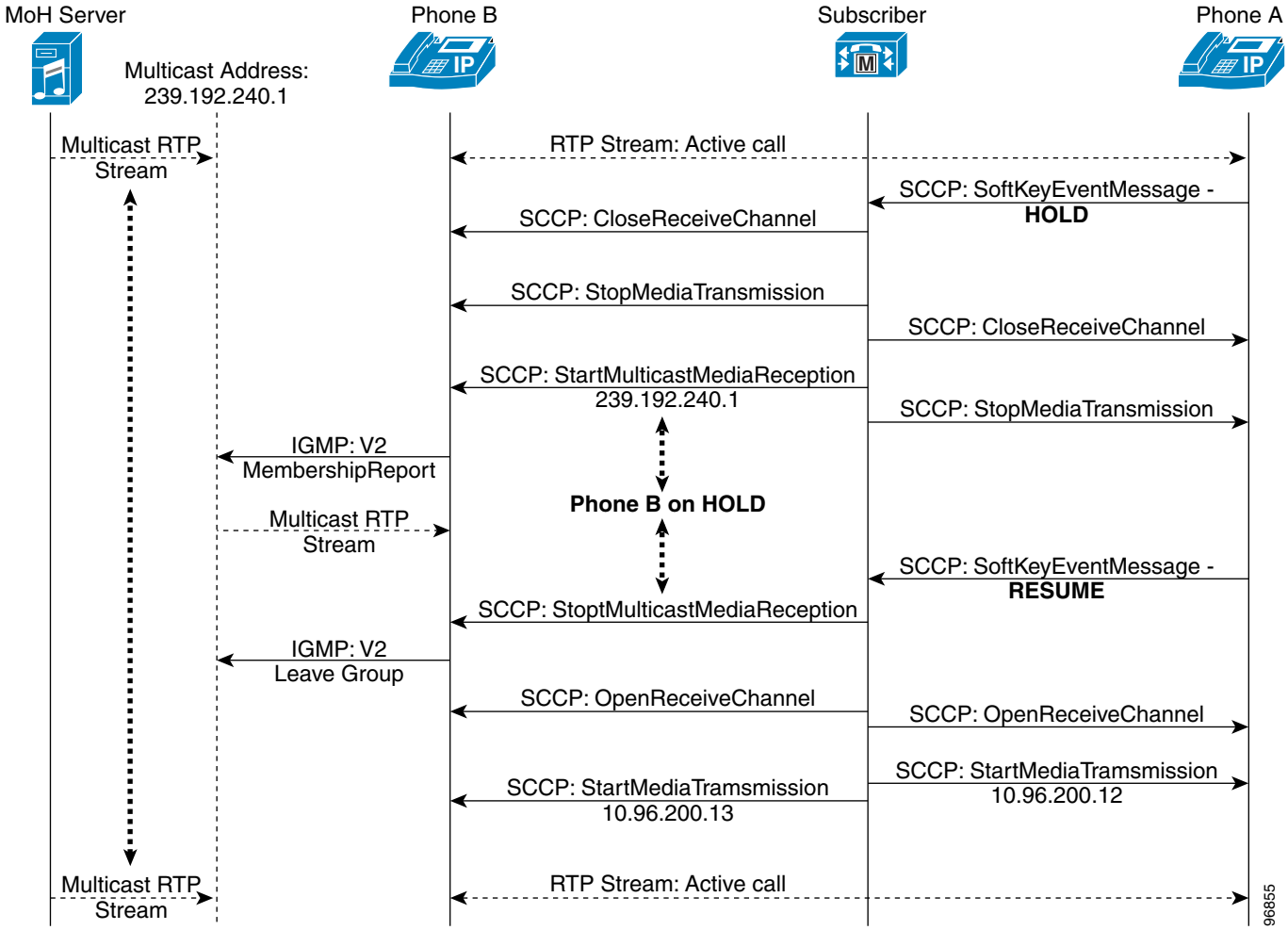
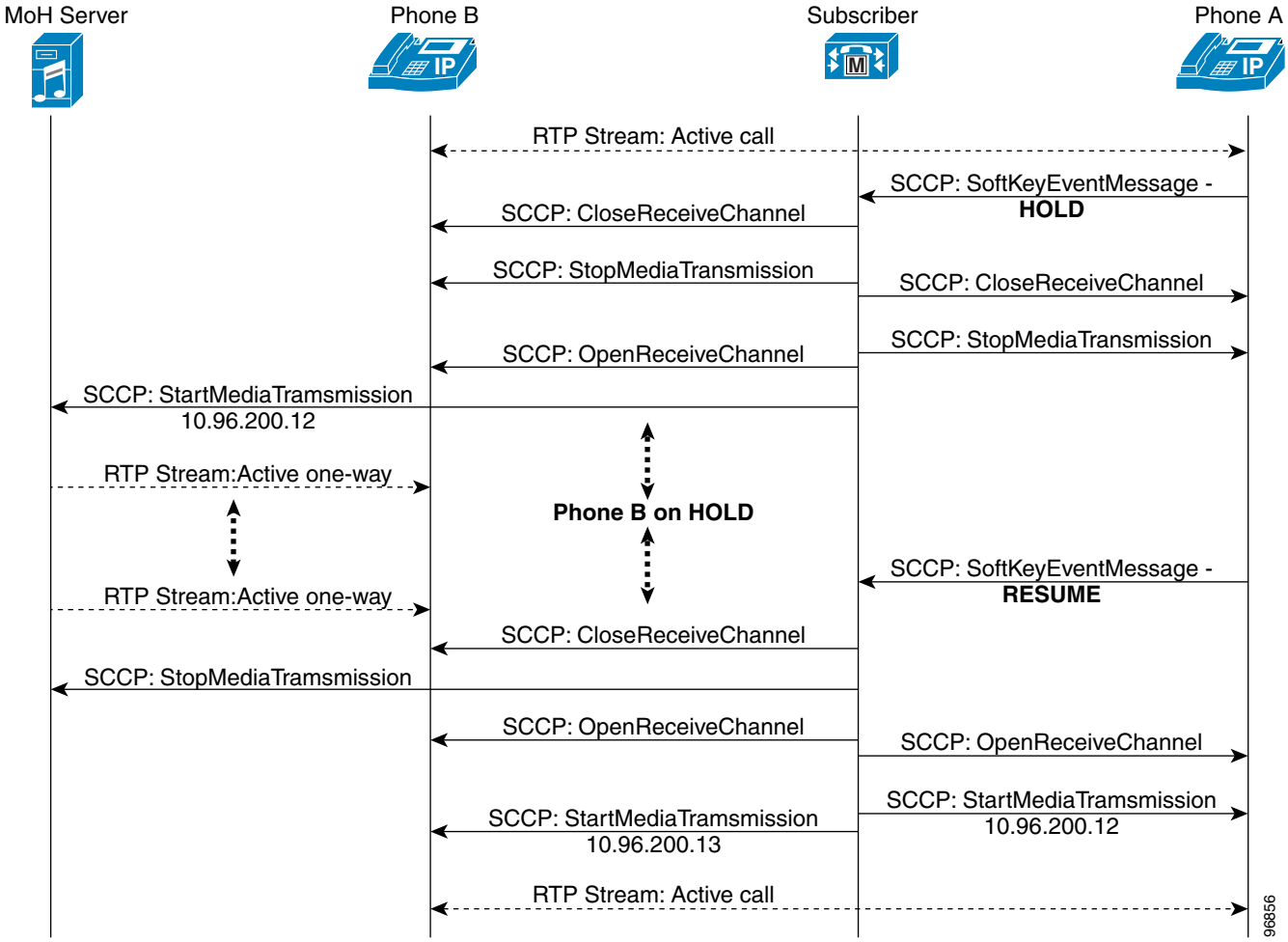


Figure 5-8 depicts a unicast MoH call flow. In this call flow diagram, when the Hold softkey is depressed at phone A, Cisco CallManager instructs both phone A and phone B to Close Receive Channel and Stop Media Transmission. This action effectively stops the RTP two-way audio stream. Up until this point, unicast and multicast MoH call flows behave exactly the same.

Next, Cisco CallManager tells phone B (the holdee) to Open Receive Channel. (This is quite different from the multicast case, where Cisco CallManager tells the holdee to Start Multicast Media Reception.) Then Cisco CallManager tells the MoH server to Start Media Transmission to the IP address of phone B. (This too is quite different behavior from the multicast MoH call flow, where the phone is prompted to join a multicast group address.) At this point, the MoH server is sending a one-way unicast RTP music stream to phone B. When phone A presses the Resume softkey, Cisco CallManager instructs the MoH

server to Stop Media Transmission and instructs phone B to Close Receive Channel, effectively ending the MoH session. As with the multicast scenario, Cisco CallManager sends a series of Open Receive Channel messages and Start Media Transmissions messages to phones A and B with each other's IP addresses. The phones are once again connected via an RTP two-way audio stream.

Figure 5-8 Detailed Unicast MoH Call Flow



Call Processing

This chapter summarizes the following design considerations related to the call processing functions of a Cisco CallManager cluster:

This chapter also contains the following section on gatekeepers, which can be used as additional call processing agents in combination with Cisco CallManager to expand or enhance the capabilities of your IP Telephony network:

Clustering Guidelines

The following guidelines apply to all Cisco CallManager clusters:

Note A cluster may contain a mix of server platforms, but all servers in the cluster must run the same Cisco CallManager software release.

- Under normal circumstances, place all members of the cluster within the same LAN or MAN.
- If the cluster spans an IP WAN, follow the specific guidelines for clustering over an IP WAN. (See Clustering Over the IP WAN, page 1-17.)
- Each server in the cluster can support a maximum of 500 H.323 (gateway and client) and digital MGCP devices.
- Cisco highly recommends that you ensure all Cisco CallManager servers are connected to the Ethernet at 100 Mbps Full Duplex. If 100 Mbps is not available on smaller deployments, then use 10 Mbps Full Duplex. You can achieve this configuration by forcing the NIC card to the required speed and duplex and by manually configuring the Ethernet switch port as well.

Note If either the server port or the Ethernet switch port is left in AUTO mode and the other port is configured manually, this will cause a mismatch. The best practice is to configure both the server port and the Ethernet switch port manually.

- Cisco recommends that you leave voice activity detection (VAD) disabled within the cluster. VAD is disabled by default in the Cisco CallManager service parameters, and you should disable it on H.323 dial peers by using the **no vad** command.

Table 6-1 lists the general types of servers you can use in a cluster, along with their main characteristics.

Table 6-1 Types of Cisco CallManager Servers

Server Type	Characteristics
Standard server (not high availability)	• Single processor • Single power supply • Non-RAID hard disk
High-availability standard server	• Single processor • Multiple power supplies • Single SCSI RAID hard disk array
High performance server	• Multiple processors • Multiple power supplies • Multiple SCSI RAID hard disk arrays

Call Processing with Cisco CallManager Releases 3.1 and 3.2

The following guidelines apply to Cisco CallManager releases 3.1 and 3.2:

- Within a cluster, you may enable a maximum of 6 servers (4 primary and 2 backup servers) with the Cisco CallManager Service. Other servers may be used for more dedicated functions such as Trivial File Transfer Protocol (TFTP), database publisher, music on hold, and so forth.
- You can configure a maximum of 800 Computer Telephony Integration (CTI) connections or associations per server, or a maximum of 3200 per cluster if they are equally balanced among the servers. For more information, refer to Computer Telephony Integration (CTI), page 12-1.
- Each H.323 device can support up to 500 H.323 calls with Cisco CallManager Release 3.1 or 1000 calls with Cisco CallManager Release 3.2.

Call Processing with Cisco CallManager Release 3.3

The following guidelines apply to Cisco CallManager release 3.3:

- Within a cluster, you may enable a maximum of 8 servers with the Cisco CallManager Service. Other servers may be used for more dedicated functions such as TFTP, publisher, music on hold, and so forth.
- You can configure a maximum of 800 CTI connections or associations per standard server, or a maximum of 3200 per cluster if they are equally balanced among the servers.

- You can configure a maximum of 2500 CTI connections or associations per MCS 7845 or equivalent server, or a maximum 10,000 per cluster if they are equally balanced among the servers.
- The maximum number of H.323 calls is limited by device weights. (See Device Weights, page 6-3.)
- Each cluster can support up to 30,000 IP phones.
- Each cluster can support a total of up to 40,000 device weight units.

Device Weights

Table 6-2 shows the base device weight for each device type, based on the consumption of memory and central processing unit (CPU) resources.

Table 6-2 Base Device Weights

Device type	Weight per Session or Voice Channel	Session or DS0 per Device	Cumulative Device Weight
IP phone	1	1	1
Analog MGCP ports	3	Varies	3 per DS0
Analog SCCP ports	1	Varies	1 per DS0
CTI route point	2	Varies	Varies[1]
CTI client port	2	1	2
CTI server port	2	1	2
CTI third-party control[2]	3	1	3
CTI agent phone[2]	6	1	6
H.323 client	3	Varies	3 per call
Intercluster trunk gateway	3	Varies	3 per call
H.323 gateway	3	Varies	3 per call
Digital MGCP T1 gateway ports	3	24	72 per T1
Digital MGCP E1 gateway ports	3	30	90 per E1
Music on hold (MoH) stream	10	20	200[3]
Transcoding resource	3	Varies	3 per session
Media Termination Point (MTP) (software)	3	24	72[4]
Conference resource (hardware)	3	Varies	3 per session
Conference resource (software)	3	24	72[4]

1. Cumulative weight of CTI route point depends on the associated CTI ports used by the application.
2. Includes the associated IP phone.
3. When MoH is installed on the same server as Cisco CallManager, the maximum number of streams is 20.
4. When installed on the same server as Cisco CallManager, the maximum number of conference sessions is 24.

BHCA Multiplier

The base weight of each device is calculated using 6 or fewer busy hour call attempts (BHCA). As the quantity of BHCA increases, so also does the number of transactions the servers are required to process and, therefore, the weight of the device on that platform. Each device requiring more than 6 BHCA has a multiplier applied to its base weight. The multiplier is calculated by dividing the BHCA by 6 and rounding up to the nearest whole number. For example, an IP phone that is making 15 BHCA would have a multiplier of 3 (15/6 = 2.5, rounded up to 3). The total device weight of the IP phone with 15 BHCA is equal to its base weight multiplied by 3. Table 6-3 illustrates the effect of the BHCA multiplier.

Note The multiplier is applied only to station or client devices and not devices that are a media resource or gateway.

Table 6-3 Effect of BHCA Multiplier on Device Weights

	0 to 6 BHCA	7 to 12 BHCA	13 to 18 BHCA	19 to 24 BHCA	25 to 30 BHCA
Multiplier	1	2	3	4	5
Device weight for example IP phone	1	2	3	4	5
Device weight for example CTI client port	2	4	6	8	10

Server Platforms

The total number of device weight units that a single Cisco CallManager can support depends on the server platform, as indicated in Table 6-4.

Table 6-4 Maximum Number of Devices per Server Platform

Server Platform Characteristics	Maximum IP Phones per Server[1]	Maximum Device Weight Units per Server	High Availability Platform[2]	High Performance Server
Cisco MCS-7845 (All models)	7500	10,000	Yes	Yes
Hewlett-Packard/Compaq DL380, Dual CPU All Cisco approved models	7500	10,000	Yes	Yes
Cisco MCS-7835 (All models) Pentium III (733-1266MHz), 1-2GB RAM	2500	5000	Yes	No
Cisco MCS-7830 Pentium III (500MHz), 1GB RAM	1500	3000	Yes	No
Cisco MCS-7830 Pentium III (500MHz), 512MB RAM	500	1000	Yes	No
Cisco MCS-7825-1133 Pentium III (1.133GHz), 1GB RAM	1000	2000	No	No
Cisco MCS-7825-800 Pentium III (800MHz), 512MB RAM	1000	2000	No	No

Table 6-4 Maximum Number of Devices per Server Platform (continued)

Server Platform Characteristics	Maximum IP Phones per Server[1]	Maximum Device Weight Units per Server	High Availability Platform[2]	High Performance Server
Cisco MCS-7822 Pentium III (550MHz), 512MB RAM	500	1000	No	No
Cisco MCS-7820 Pentium III (500MHz), 512MB RAM	500	1000	No	No
Cisco SPE 310 (ICS 7750) Pentium III (700MHz), 512MB RAM	1000	2000	No	No
Hewlett-Packard/Compaq DL380 All Cisco approved models	2500	5000	Yes	No
Hewlett-Packard/Compaq DL320 All Cisco approved models	1000	2000	No	No
IBM xSeries 34*x* All Cisco approved models	2500	5000	Yes	No
IBM xSeries 33*x* All Cisco approved models	1000	2000	No	No
Cisco MCS-7815 Celeron (1GHz), 512MB RAM	200	400	No	No

1. A platform that is not a high availability server can support a maximum of 500 IP phones in a non-redundant installation.
2. A high availability server supports redundancy for both the power supplies and the hard disks.

For the latest information on supported platforms and specific hardware configurations, refer to the online documentation at

http://www.cisco.com/en/US/products/hw/voiceapp/ps378/prod_brochure_list.html

Dial Plan Weights

Cisco CallManager uses device weights to control the number of physical devices that may register with any subscriber server. The BHCA multiplier adjusts the device weight of devices that have a higher impact on a subscriber as they make or receive more calls in the busy hour.

Dial plan weights control the number of dial plan parameters that you can configure on a Cisco CallManager server. Subscriber dial plan weights relate to individual device configuration parameters (specifically, line appearances), and global dial plan weights relate to global configuration parameters (such as route patterns and translations patterns) that are not associated with a particular device.

Subscriber dial plan weights consist of the following main types:

- *IP phone weight* is associated with the subscriber server to which the phone (or other dialable device) is normally registered. The dial plan weight of an IP phone is independent of its device weight. (See Device Weights, page 6-3.)
- *Line weight* (unique or shared appearance) is associated with the subscriber server to which the device is normally registered, as defined by the Cisco CallManager redundancy groups.
- *Reachability weight* is associated with all other subscribers in the cluster that have to be able to reach (call) a given line.

Global dial plan weights consist of route pattern and translation pattern weights.

A line appearance is defined as any device (IP phone, CTI port, CTI route point, or H.323 client) that has a directory number (DN) assigned to it. Every additional line appearance or DN on that same device also has an associated dial plan weight.

In summary, dial plan components contribute the following weights:

- Subscriber dial plan weights
  - IP phone or other dialable device (excluding line appearances) = 5
  - Unique line appearance = 5
  - Shared line appearance = 4
  - Reachability per line appearance = 3
- Global dial plan weights
  - Route pattern = 2
  - Translation pattern = 1

For example, if an IP phone with a unique line appearance (such as extension 1000) is registered with Subscriber A, the dial plan weight on Subscriber A is 10 (5 for the IP phone and 5 for the unique line appearance). All other subscribers in the cluster also need information so that they can reach extension 1000 on Subscriber A, and they have a dial plan reachability weight of 3 each. Each additional unique line appearance on the same phone would add 5 dial plan weight units to Subscriber A and 3 units to each of the other subscribers in the cluster, while each shared line appearance on the same phone would add 4 dial plan weight units to Subscriber A and 3 units to each of the other subscribers in the cluster.

Global dial plan weights apply to all subscriber servers in the cluster. For example, a route pattern of 9.911 adds a dial plan weight of 2 to each subscriber server in the cluster.

Table 6-5 lists guidelines for the amount of physical memory to install on a subscriber server, based on the number of phones, number of line appearances, and complexity of the dial plan configured on that server.

Table 6-5 Server Memory Requirements for Dial Plan Weights

Total Dial Plan Weight Units on Subscriber Server	Server Memory Requirements
Up to 15,000	512 MB of RAM installed
Up to 35,000	768 MB of RAM installed
Up to 70,000	1 GB of RAM installed.
Up to 140,000	2 GB of RAM installed

For normal deployments, do not configure more than 140,000 dial plan weight units on each subscriber server.

Note Dial plan weight calculations are separate from device weight calculations. The subscriber servers and the cluster must comply with both the device weight limit and the dial plan weight limit. If either limit is exceeded, reduce the number of devices, line appearances, and so on, to achieve compliance.

A Cisco CallManager cluster with a very large dial plan containing many gateways, route patterns, translation patterns, and partitions can take an extended amount of time to initialize when the Cisco CallManager Service is first started. If the system does not initialize within the default time, there are service parameters that you can increased to allow additional time for the configuration to initialize. For details on the service parameters, refer to the Cisco CallManager *Administration Guide* and *System Guide*, available on Cisco.com.

Call Processing Redundancy

With all versions of Cisco CallManager, you can choose from the following redundancy configurations:

- Two to one (2:1) — For every two primary subscribers, there is one shared backup subscriber.
- One to one (1:1) — For every primary subscriber, there is a backup subscriber.

The 1:1 redundancy scheme allows upgrades with only the failover periods impacting the cluster. The failover mechanism has been enhanced so that you can achieve failover rates for the IP phones of approximately 100 registrations per second.

Cisco CallManager Release 3.3 supports up to eight subscribers (servers with the Cisco CallManager service enabled), so you may have as many as four primary and four backup subscribers in a cluster.

The 1:1 redundancy scheme enables you to upgrade the cluster using the following method.

Step 1 Upgrade the publisher server.

Step 2 Upgrade dedicated TFTP and music on hold (MoH) servers.

Step 3 Upgrade all backup subscribers. This step will impact some users if 50/50 load balancing is implemented.

Step 4 Fail-over the primary subscribers to their backups, and stop the Cisco CallManager service on the primaries.

Step 5 Upgrade the primaries, then re-enable the Cisco CallManager service.

With this upgrade method, there is no period (except for the failover period) when devices are registered to subscriber servers that are running different versions of the Cisco CallManager software. This factor can be important because the Intra-Cluster Communication Signaling (ICCS) protocol that communicates between subscribers can detect a different software version and shut down communications to that subscriber. This action could potentially partition a cluster for call processing, but SQL and LDAP replication would not be affected.

The 2:1 redundancy scheme allows for fewer servers in a cluster, but it can potentially result in an outage during upgrades.

Note You must use 1:1 redundancy when 10,000 or more IP phones are registered on the two primary subscribers because there cannot be more than 10,000 backup registrations on a single backup subscriber.

Cluster Configurations for Redundancy

The following figures illustrate typical cluster configurations to provide call processing redundancy with Cisco CallManager.

Figure 6-1 Basic Redundancy Schemes

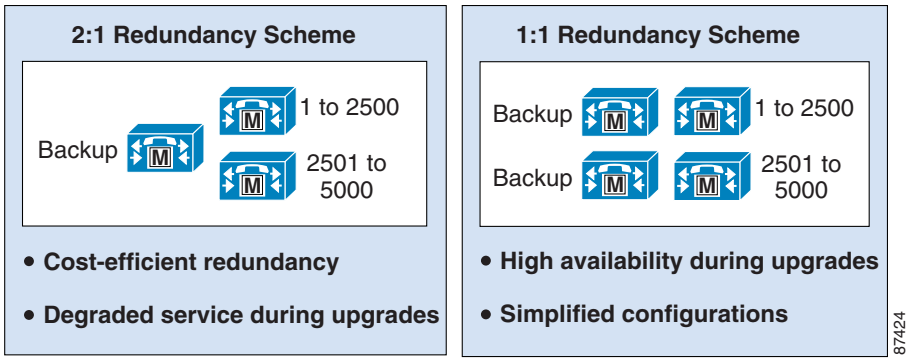


Figure 6-2 2:1 Redundancy with Cisco CallManager Release 3.x on Standard Server

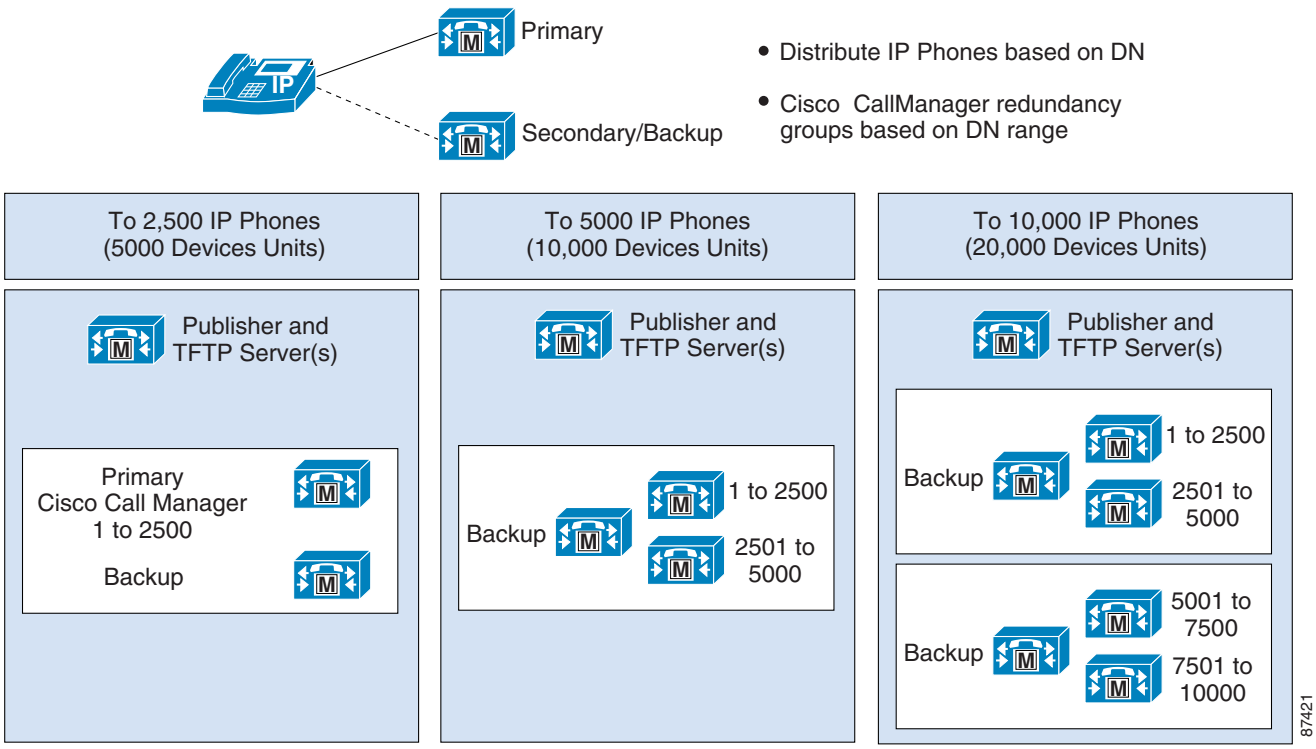


Figure 6-3 1:1 Redundancy with Cisco CallManager Release 3.3 on Standard Server with 50/50 Load Balancing

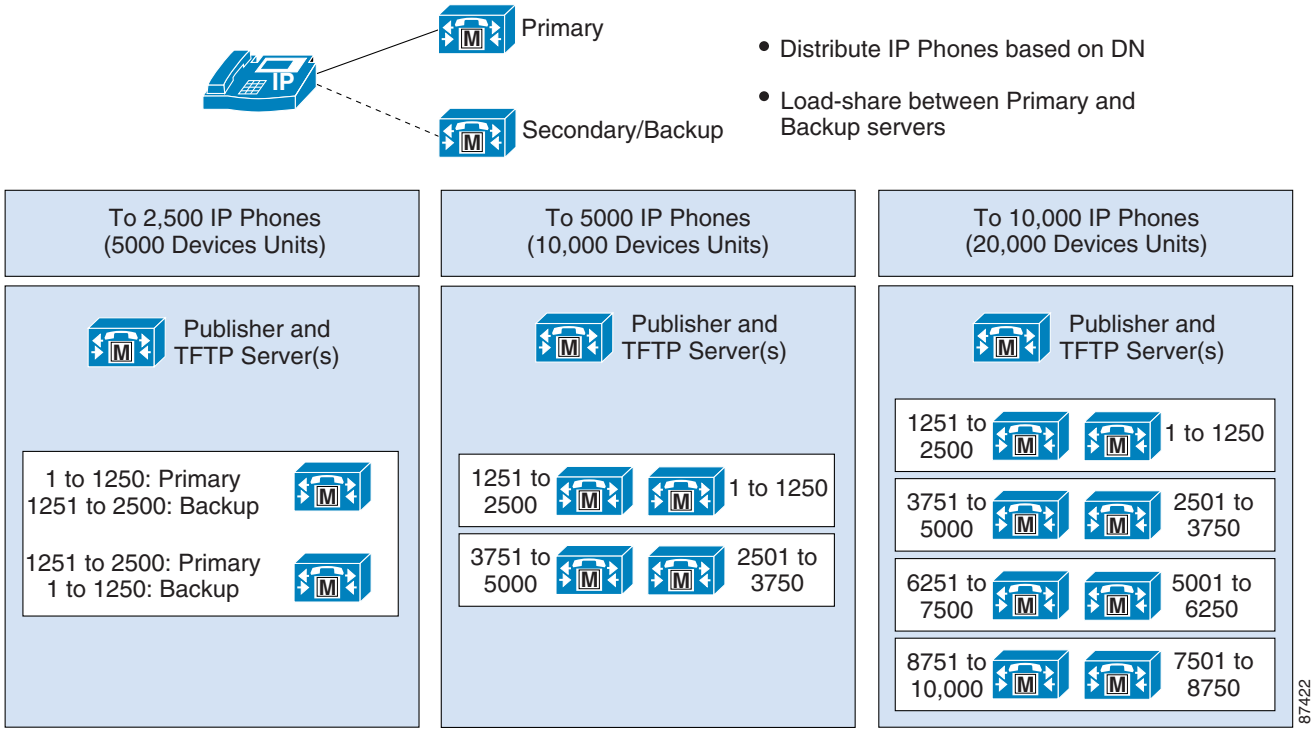
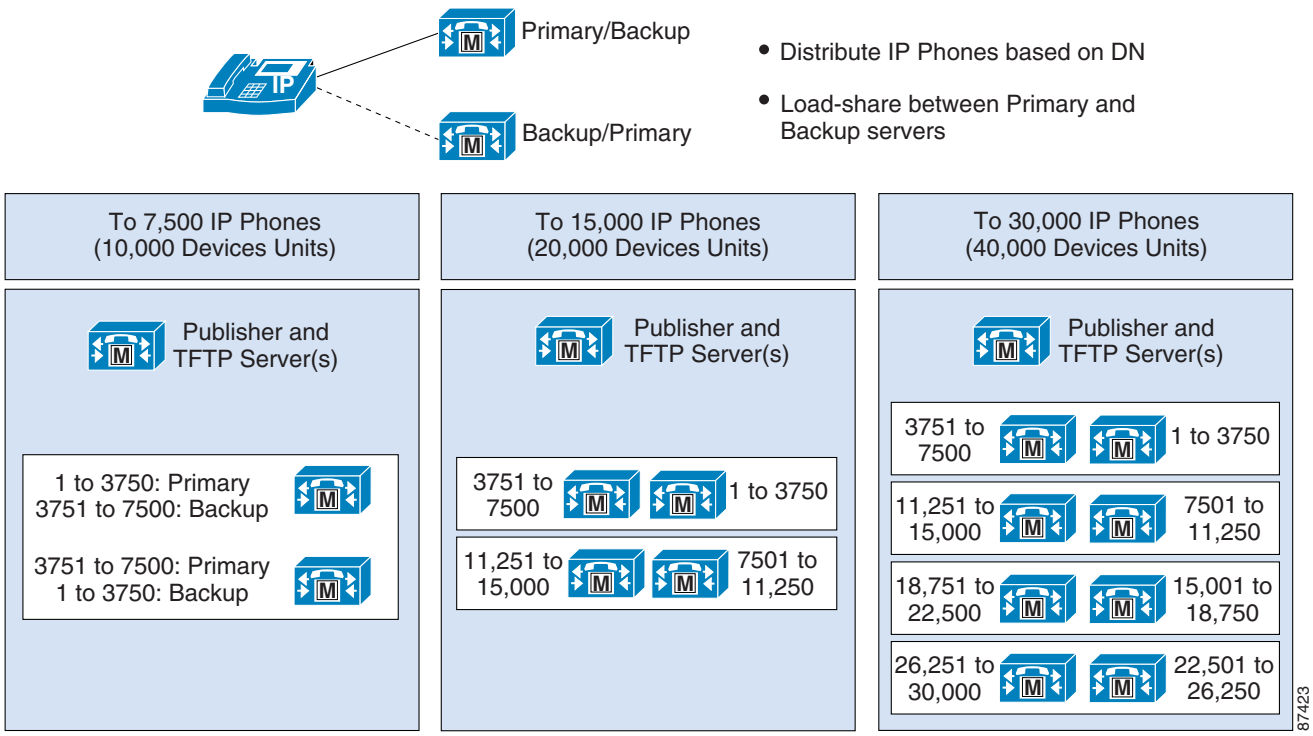


Figure 6-4 Redundancy with Cisco CallManager Release 3.3 on MCS-7845 High Performance Server with 50/50 Load Balancing



Load Balancing

An additional benefit of using the 1:1 redundancy scheme is that it enables you to balance the devices over the primary and backup server pairs. Normally a backup server has no devices registered unless its primary is unavailable.

With load balancing, you can move up to half of the device load from the primary to the secondary subscriber by using the Cisco CallManager redundancy groups and device pool settings. In this way, you can reduce by 50% the impact of any server becoming unavailable.

To plan for 50/50 load balancing, stay within the device weight limits for a single server to allow for the failover scenario. For example, MCS 7835 servers have a total server limit of 5000 device weight units and 2500 IP phones. In a 1:1 redundancy pair, you can split the load between the two subscribers, configuring each subscriber with 2500 device weight units and 1250 IP phones. (See the configuration for 2500 IP phones in Figure 6-3.)

To provide for failover conditions when only one server is active, make sure that total device weight units, IP phones, CTI limits, and so on, for the redundant pair do not exceed the limits allowed for a single server.

Secondary TFTP Server

For additional load balancing and redundancy, you can install a second Trivial File Transfer Protocol (TFTP) server in the cluster. The TFTP server facilitates the downloading of configuration files, device loads (operating code), and ring types. You should configure the secondary server for TFTP only, with the Cisco CallManager service turned off so that no phones can register with that server. Once the Cisco CallManager service is running on another server in that cluster, the intra-cluster communications will populate the database, and all of the phone configurations will be created and placed into memory on the TFTP servers.

Pay special attention to the secondary TFTP server if you have to change the phone load for a specific type of phone in the cluster, but you are not upgrading the Cisco CallManager software at the same time. In that case, be sure to copy the new phone load manually to each TFTP server.

With Cisco CallManager Release 3.3 and later, phone configuration files are not stored by default on the hard drive of the TFTP server, as in earlier versions of Cisco CallManager. By default, all phone configuration files are created and placed into RAM on the TFTP servers. You can change this default setting to place the phone configuration files on the hard drive of the TFTP server, but doing so will impact TFTP performance. Therefore, Cisco recommends that you do not change this default setting.

Gatekeeper Considerations

A Cisco IOS gatekeeper can provide call routing and call admission control for up to 100 Cisco CallManager clusters in a distributed call processing environment. You can also implement a hybrid Cisco CallManager and toll bypass network by using Cisco IOS gatekeepers to provide communication and call admission control between the H.323 gateways and Cisco CallManager.

Gatekeeper call admission control is a policy-based scheme requiring static configuration of available resources. The gatekeeper is not aware of the network topology, so it is limited to hub-and-spoke topologies.

The Cisco 2600, 3600, 3700, and 7200 Series routers all support the gatekeeper feature. The Cisco IOS gatekeeper feature is available in both the IP/H323 and Enterprise Multimedia Conference Manager (MCM) feature sets. Choose the correct feature set based on your deployment requirements, as listed in Table 6-6.

Table 6-6 Supported Gatekeeper Features

Feature Set	Voice Gateway	Proxy	Multi-Protocol Support	Alternate Gatekeeper (Clustering)	Alternate Endpoint (EP)	Hot Standby Router Protocol (HSRP)
IP/H323	No	Yes	No (IP only)	Yes, with Cisco IOS Release 12.2(13)T or later	Yes	Yes
Enterprise MCM	Yes	Yes	Yes	Yes, with Cisco IOS Release 12.2(13)T or later	Yes	Yes

You can configure Cisco IOS gatekeepers in a number of different ways for redundancy, load balancing, and hierarchical call routing. To accomplish these goals, it is helpful to understand the parsing logic in the gatekeeper. Figure 6-5 illustrates the parsing logic for an Admission Request (ARQ), and Figure 6-6 illustrates the parsing logic for a Location Request (LRQ).

To initiate a call, an endpoint sends an Admission Request (ARQ) to the gatekeeper. The ARQ contains either an H.323 ID or the E.164 address of the destination, or called party, as well as the E.164 address or H.323 ID of the source, or calling party.

The ARQ also contains the requested call bandwidth, which should be the upper limit of both the transmitted and received bit rates for all video and voice channels. The bandwidth does not allow for any transport, IP, or RTP overhead.

If the ARQ contains the E.164 address (with Cisco CallManager, the ARQ always will contain an E.164 address), the ARQ may or may not contain the technology prefix. If the ARQ does not contain the technology prefix, the gatekeeper uses the default technology prefix if one is configured. See the **gw-type-prefix** command in the section on Centralized Gatekeeper Configuration, page 6-14.

The gatekeeper responds to the ARQ with an Admission Confirm (ACF) if the requested bandwidth is available and the called endpoint is registered with the gatekeeper. The ACF will contain the IP address of the destination endpoint. If the bandwidth is unavailable or the called endpoint is not registered, the gatekeeper returns an Admission Reject (ARJ) to the calling endpoint.

Upon receipt of an ACF from the gatekeeper, the source endpoint can send a setup message directly to the destination endpoint by using the IP address returned in the ACF.

Figure 6-5 Gatekeeper Address Resolution for an ARQ

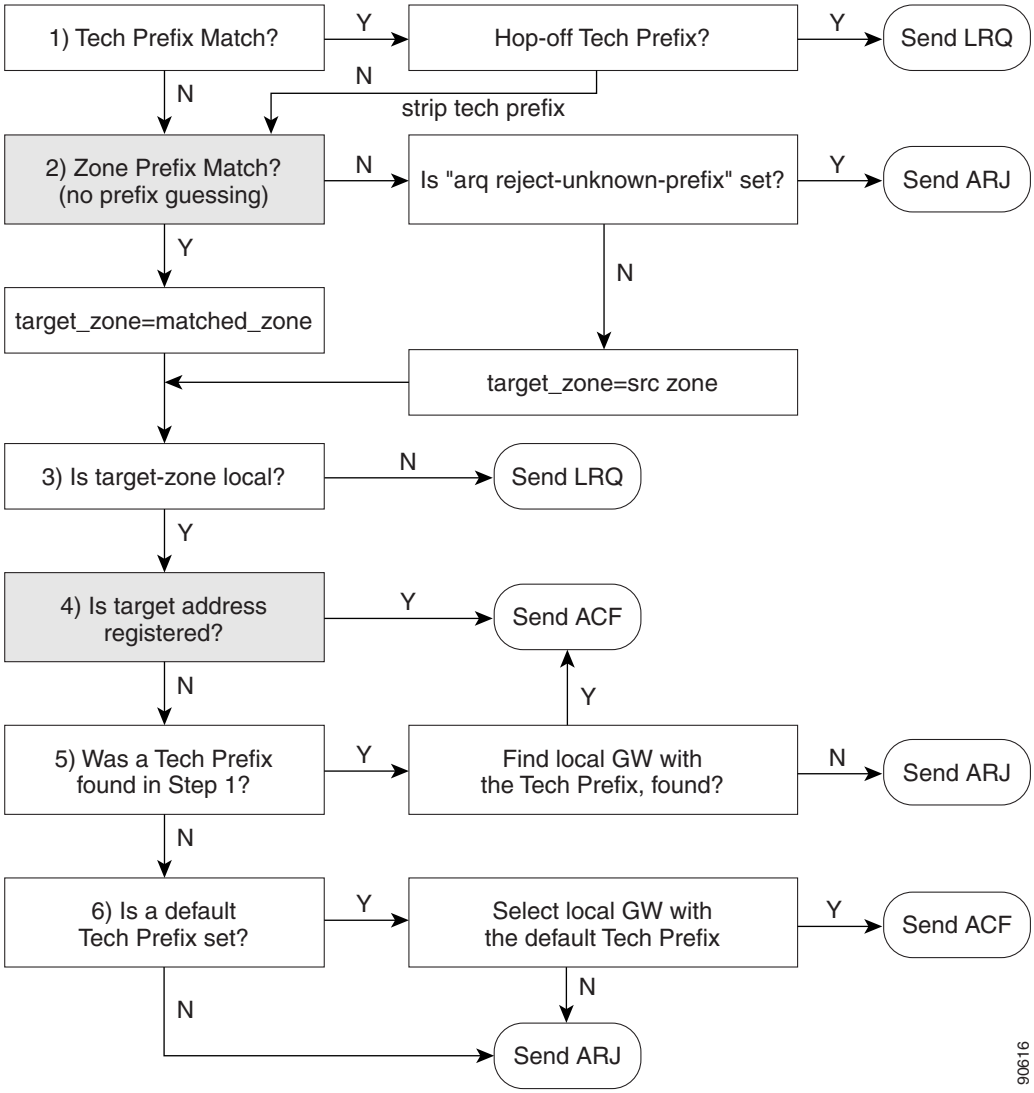
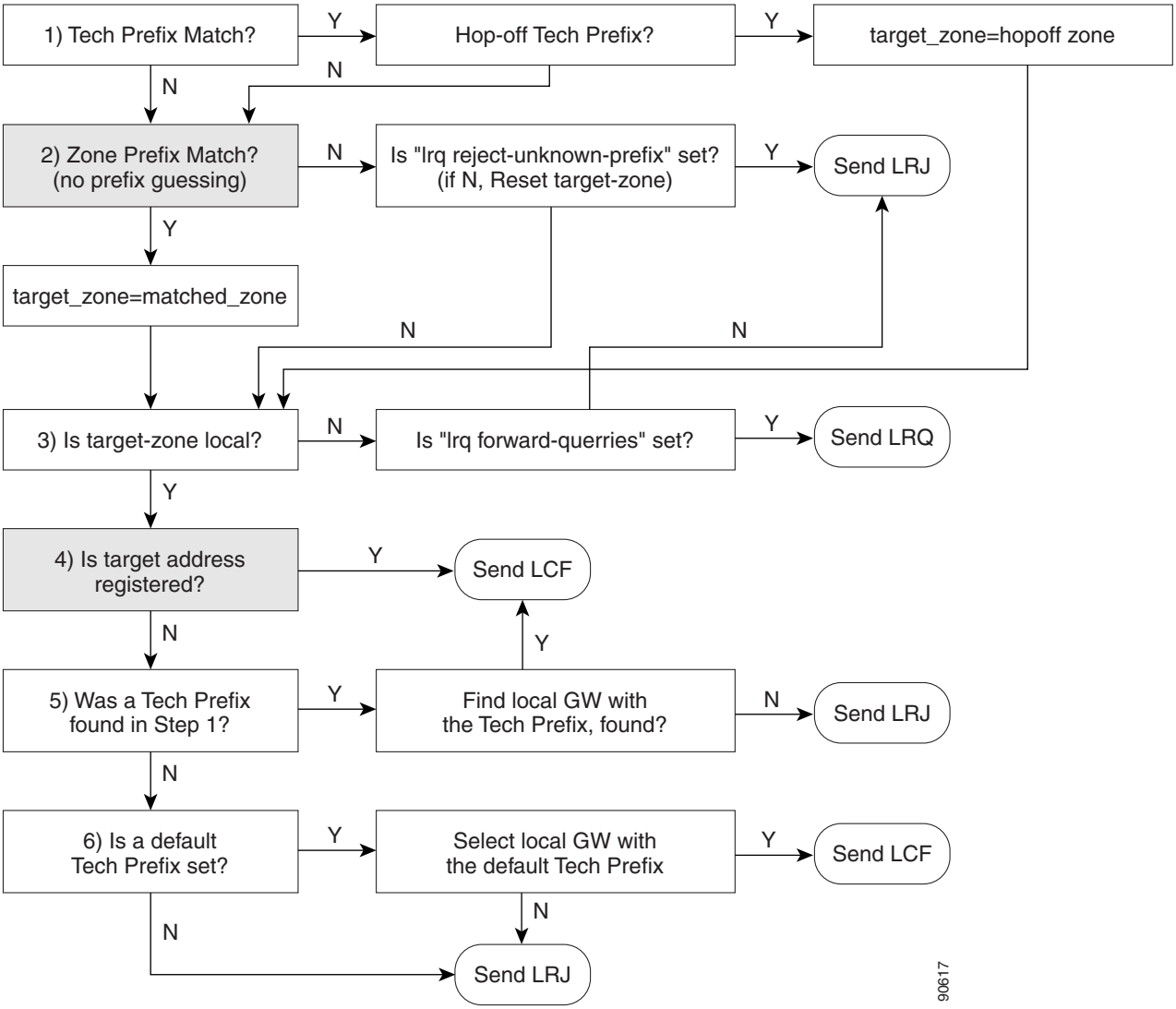


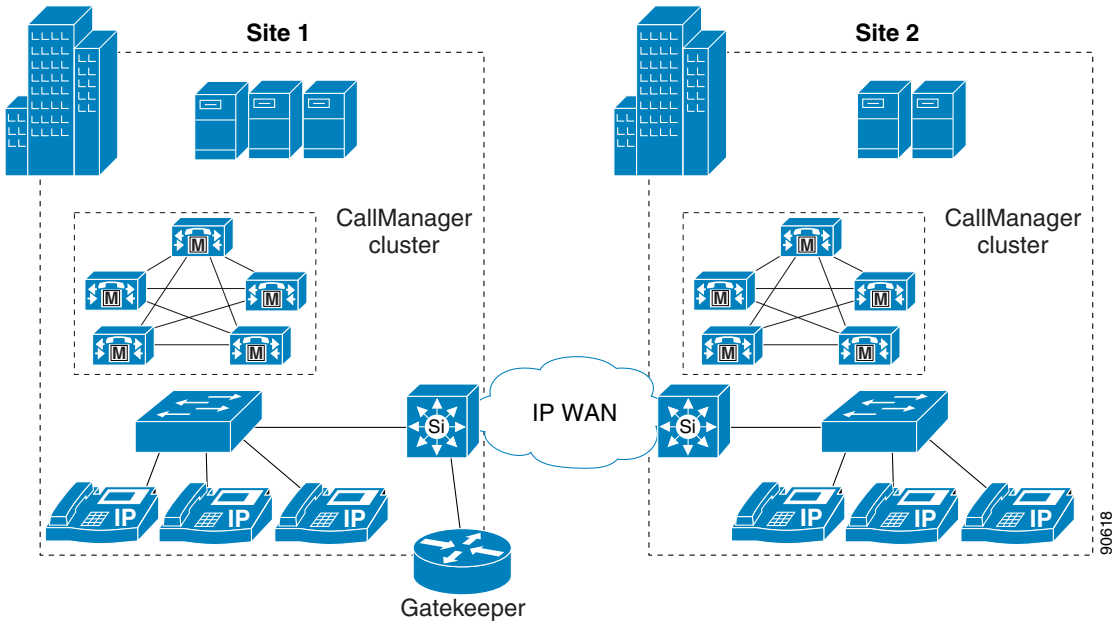
Figure 6-6 Gatekeeper Address Resolution for an LRQ



Centralized Gatekeeper Configuration

A single gatekeeper can support call routing between clusters and call admission control for up to 100 Cisco CallManager clusters. Figure 6-7 illustrates a distributed call processing environment with two Cisco CallManager clusters and a single, centralized gatekeeper.

Figure 6-7 Centralized Gatekeeper Supporting Two Clusters



Example 6-1 shows the gatekeeper configuration for the example in Figure 6-7.

Example 6-1 Configuration for Centralized Gatekeeper

```
gatekeeper
 zone local GK-Site1 customer.com 10.1.10.100
 zone local GK-Site2  customer.com
 zone prefix GK-Site1 408.......
 zone prefix GK-Site2 212.......
 bandwidth interzone GK-Site1 160
 bandwidth interzone GK-Site2 160
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

The following notes also apply to Figure 6-7:

- Each Cisco CallManager cluster has a local zone configured to support Cisco CallManager trunk registrations.
- A zone prefix is configured for each zone to allow inter-zone and inter-cluster call routing.
- The "." (dot) indicates a wildcard match of a single digit. In this example, an E.l64 address has to be 10 digits long and start with either 408 or 212. The * (star) can also be used as a wildcard to indicate any number of digits. If the zone prefixes in Example 6-1 where 408* and 212*, then a number of any length, starting with either 408 or 212, would be routed correctly.

- Bandwidth statements are configured for each site. Cisco recommends that you use the **bandwidth interzone** command. Using the **bandwidth total** command can cause issues in some configurations. Bandwidth is measured in kilobits per second (kbps).
- The **gw-type-prefix 1# default-technology** command allows all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.

  Technology prefixes indicate the type of call being made.The specific values used as technology prefixes are arbitrary and are defined by the network administrator. The same values should be used consistently throughout the entire deployment.

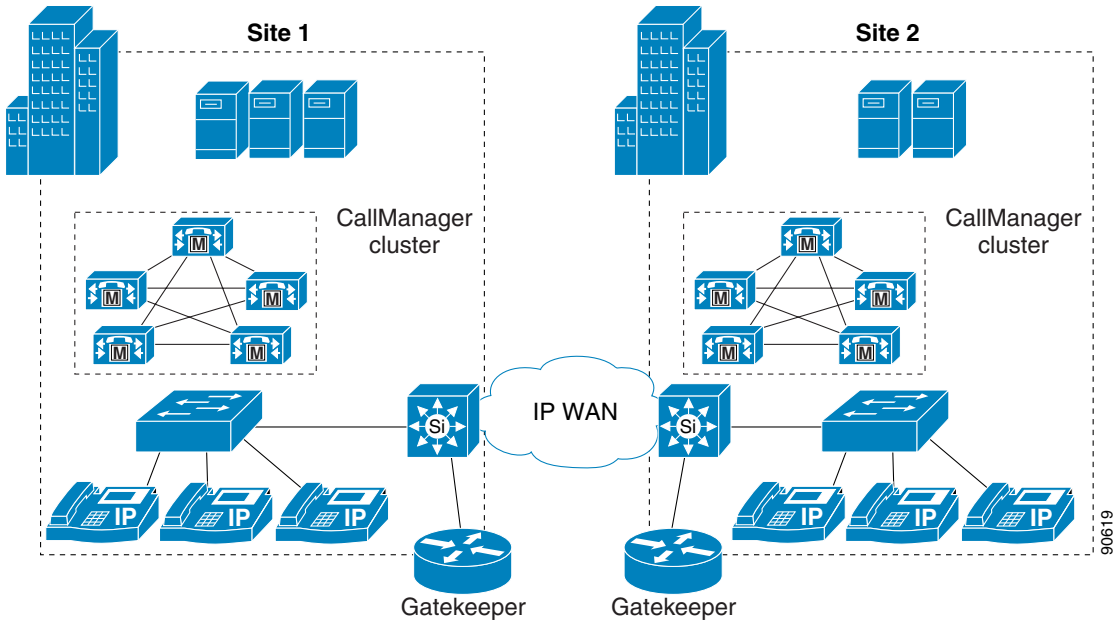
  Technology prefixes are sent as a prefix to the E.164 address (phone number) to indicate whether the call is voice, video, or some other type. The # symbol is generally used to separate the prefix from the E.164 number. If a prefix is not included, the default technology prefix is used to route the call. There can be only one default technology prefix for the deployment.

  Cisco IOS gateways automatically add a technology prefix to outbound calls if the gateway has a prefix configured. The gateway also automatically strips the prefix from incoming H.323 calls. Cisco CallManager does not automatically send a technology prefix even if one is configured on the trunk. You can, however, configure a prefix on the Cisco CallManager H.323 Trunk Gateway configuration pages. You must also strip the technology prefix from inbound calls to Cisco CallManager by using either translation patterns or significant digits on the H.323 trunk.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Cisco CallManager trunks.

Distributed Gatekeeper Configuration

Gatekeepers can be distributed to conserve bandwidth or to provide local call routing for H.323 gateways in case of a WAN failure. Figure 6-8 illustrates a distributed call processing environment with two clusters and two gatekeepers.

Figure 6-8 Distributed Gatekeepers Supporting Two Clusters



Example 6-2 shows the gatekeeper configuration for Site 1 in Figure 6-8.

Example 6-2 Gatekeeper Configuration for Site 1

```
gatekeeper
 zone local GK-Site1 customer.com 10.1.10.100
 zone remote GK-Site2  customer.com 10.1.11.100
 zone prefix GK-Site1 408.......
 zone prefix GK-Site2 212.......
 bandwidth remote 160
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

The following notes apply to Example 6-2:

- A local zone is configured for registration of local Cisco CallManager cluster trunks.
- A remote zone is configured for routing calls to the Site 2 gatekeeper.
- Zone prefixes are configured for both zones for inter-zone call routing.
- The **bandwidth remote** command is used to limit bandwidth between the local zone and any other remote zone.
- The **gw-type-prefix 1# default-technology** command allows all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Cisco CallManager trunks.

Example 6-3 shows the gatekeeper configuration for Site 2 in Figure 6-8.

Example 6-3 Gatekeeper Configuration for Site 2

```
gatekeeper
 zone local GK-Site2 customer.com 10.1.11.100
 zone remote GK-Site1  customer.com 10.1.10.100
 zone prefix GK-Site2 212.......
 zone prefix GK-Site1 408.......
 bandwidth remote 160
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

The following notes apply to Example 6-3:

- A local zone is configured for registration of local Cisco CallManager cluster trunks.
- A remote zone is configured for routing calls to the Site 1 gatekeeper.
- Zone prefixes are configured for both zones for inter-zone call routing.
- The **bandwidth remote** command is used to limit bandwidth between the local zone and any other remote zone.
- The **gw-type-prefix 1# default-technology** command allows all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Cisco CallManager trunks.

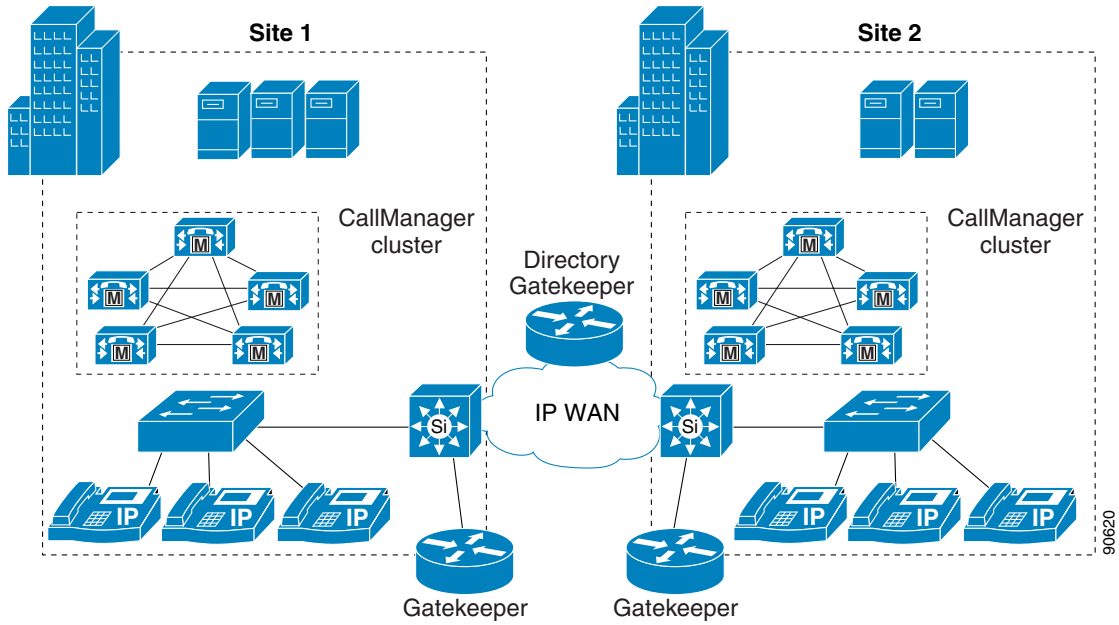
Distributed Gatekeeper Configuration with Directory Gatekeeper

Because there is no gatekeeper protocol available to update gatekeeper routing tables, use of a directory gatekeeper can help make distributed gatekeeper configurations more scalable and more manageable. Implementing a directory gatekeeper makes gatekeeper configurations at each site simpler and moves most of the configuration for inter-zone communication into the directory gatekeeper.

Without a directory gatekeeper, you would have to add an entry in every gatekeeper on the network every time you add a new zone on one of the gatekeepers. However, with a directory gatekeeper, you can add the new zone in the local gatekeeper and the directory gatekeeper only. If the local gatekeeper cannot resolve a call request locally, it forwards that request to the directory gatekeeper with a matching zone prefix.

Figure 6-9 illustrates a Cisco CallManager distributed call processing environment with distributed gatekeepers for local call routing and a directory gatekeeper to provide call routing between gatekeepers.

Figure 6-9 Distributed Gatekeepers with a Directory Gatekeeper



Example 6-4 shows the gatekeeper configuration for Site 1 in Figure 6-9. Because the Site 1 and Site 2 gatekeeper configurations are almost identical in this example, only Site 1 is covered here.

Example 6-4 Gatekeeper Configuration for Site 1, with Directory Gatekeeper

```
gatekeeper
 zone local GK-Site1 customer.com 10.1.10.100
 zone remote DGK  customer.com 10.1.10.101
 zone prefix GK-Site1 408.......
 zone prefix DGK ..........
 bandwidth remote 160
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

The following notes also apply to Example 6-4:

- A local zone is configured for registration of local Cisco CallManager cluster trunks.
- A remote zone is configured for the directory gatekeeper.
- Zone prefixes are configured for both zones for inter-zone call routing.
- The directory gatekeeper zone prefix is configured with 10 dots. This pattern matches any unresolved 10-digit dial strings. Multiple zone prefixes can be configured for a single zone, allowing matches on different length dial strings. A wildcard (*) can also be used for a directory gatekeeper zone prefix, but this method can cause call routing issues in some instances.
- The **bandwidth remote** command is used to limit bandwidth between the local zone and any other remote zone.
- The **gw-type-prefix 1# default-technology** command allows all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Cisco CallManager trunks.

Example 6-5 shows the directory gatekeeper configuration for the example in Figure 6-9.

Example 6-5 Directory Gatekeeper Configuration

```
gatekeeper
 zone local DGK customer.com 10.1.10.101
 zone remote GK-Site1 customer.com 10.1.10.100
 zone remote GK-Site2 customer.com 10.1.11.100
 zone prefix GK-Site1 408*
 zone prefix GK-Site2 212*
 lrq forward-queries
 no shutdown
```

The following notes also apply to Example 6-5:

- A local zone is configured for the directory gatekeeper.
- Remote zones are configured for the each remote gatekeeper.
- Zone prefixes are configured for both remote zones for inter-zone call routing. The wildcard (*) is used in the zone prefix to simplify configuration. Calls will not be routed to the DGK zone, so no prefix is required for it.
- The **lrq forward-queries** command allows the directory gatekeeper to forward an LRQ received from another gatekeeper.

Gatekeeper Redundancy

With gatekeepers providing all call routing and admission control for intercluster communication, redundancy is required. Prior to Cisco CallManager Release 3.3, the only method for providing gatekeeper redundancy was Hot Standby Router Protocol (HSRP); beginning with Cisco CallManager Release 3.3, gatekeeper clustering and redundant gatekeeper trunks are also available as methods of providing gatekeeper redundancy. The following sections describe these methods.

Note Cisco recommends that you use gatekeeper clustering to provide gatekeeper redundancy whenever possible. Use HSRP for redundancy only if gatekeeper clustering is not available in your software feature set. (See Table 6-6 for availability of gatekeeper clustering.)

Hot Standby Router Protocol (HSRP)

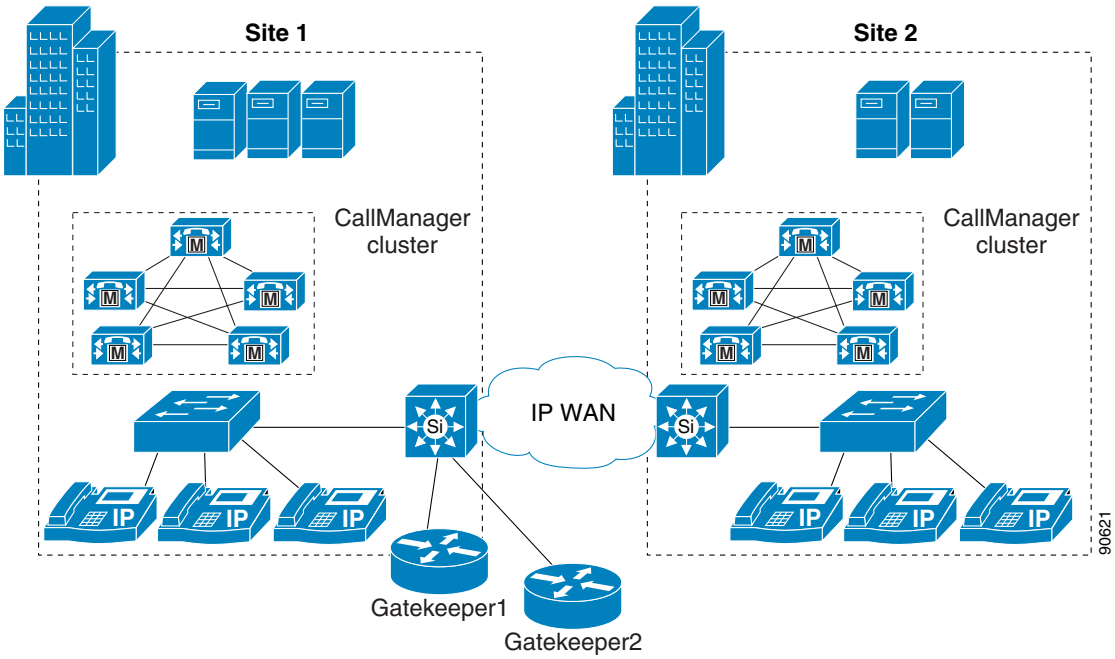
Hot Standby Router Protocol (HSRP) is the only option for gatekeeper redundancy with Cisco CallManager prior to Release 3.3. HSRP does not provide the features required to build a redundant and scalable gatekeeper network, therefore it should be used only in Cisco CallManager environments prior to Release 3.3.

The following guidelines apply to HSRP:

- Only one gatekeeper is active at a time.
  - The standby gatekeeper does not process any calls unless the primary fails.
  - No load balancing features are available.
- All gatekeepers must reside in the same subnet or location.
- No state information is available during failover.
- After a failover, the standby gatekeeper is not aware of the calls that are already active, so over-subscription of the bandwidth is possible.
- Failover time is substantial because the endpoints have to re-register with the HSRP standby gatekeeper before calls can be placed. The failover time is dependant on the settings of the registration timers.

Figure 6-10 show a network configuration using HSRP for gatekeeper redundancy.

Figure 6-10 Gatekeeper Redundancy Using HSRP



Example 6-6 shows the configuration for Gatekeeper 1, and Example 6-7 shows the configuration for Gatekeeper 2 in Figure 6-10. Both configurations are identical except for the HSRP configuration on the Ethernet interface.

Example 6-6 Configuration for Gatekeeper 1

```
interface Ethernet0/0
 ip address 10.1.10.100 255.255.255.0
 half-duplex
 standby 10 ip
 standby 10 ip 10.1.10.110

gatekeeper
 zone local GK-Site1 customer.com 10.1.10.100
 zone local GK-Site2  customer.com
 zone prefix GK-Site1 408.......
 zone prefix GK-Site2 212.......
 bandwidth interzone default 160
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

Example 6-7 Configuration for Gatekeeper 2

```
interface Ethernet0/0
 ip address 10.1.10.101 255.255.255.0
 half-duplex
 standby 10 ip
 standby 10 ip 10.1.10.110 secondary

gatekeeper
 zone local GK-Site1 customer.com 10.1.10.101
 zone local GK-Site2 customer.com
 zone prefix GK-Site1 408.......
 zone prefix GK-Site2 212.......
 bandwidth interzone default 160
 gw-type-prefix 1#* default-technology
 arq reject-unknown-prefix
 no shutdown
```

The following notes also apply to Example 6-6 and Example 6-7:

- Each router has **standby** commands configured for HSRP and to identify the virtual IP address used by each. Gatekeeper 2 is configured as the backup with the command **standby 10 ip 10.1.10.110 secondary**, where the keyword **secondary** defines this router as the backup.
- Each Cisco CallManager cluster has a local zone configured on each router to support Cisco CallManager trunk registrations.
- A zone prefix is configured for each zone on both routers, allowing inter-zone and inter-cluster call routing.
- Bandwidth statements are configured on each router for both sites. Cisco recommends that you use the **bandwidth interzone** command. Using the **bandwidth total** command can cause issues in some configurations.
- The **gw-type-prefix 1# default-technology** command is configured on both routers, allowing all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.

- The **arq reject-unknown-prefix** command is configured on both routers to guard against potential call routing loops across redundant Cisco CallManager trunks.

Gatekeeper Clustering (Alternate-Gatekeeper)

Gatekeeper clustering (Alternate-Gatekeeper) enables the configuration of a "local" gatekeeper cluster, with each gatekeeper acting as primary for some Cisco CallManager trunks and an alternate for others. Gatekeeper Update Protocol (GUP) is used to exchange state information between gatekeepers in a local cluster. GUP tracks and reports CPU utilization, memory usage, active calls, and number of registered endpoints for each gatekeeper in the cluster. Load balancing is supported by setting thresholds for any of the following parameters in the GUP messaging:

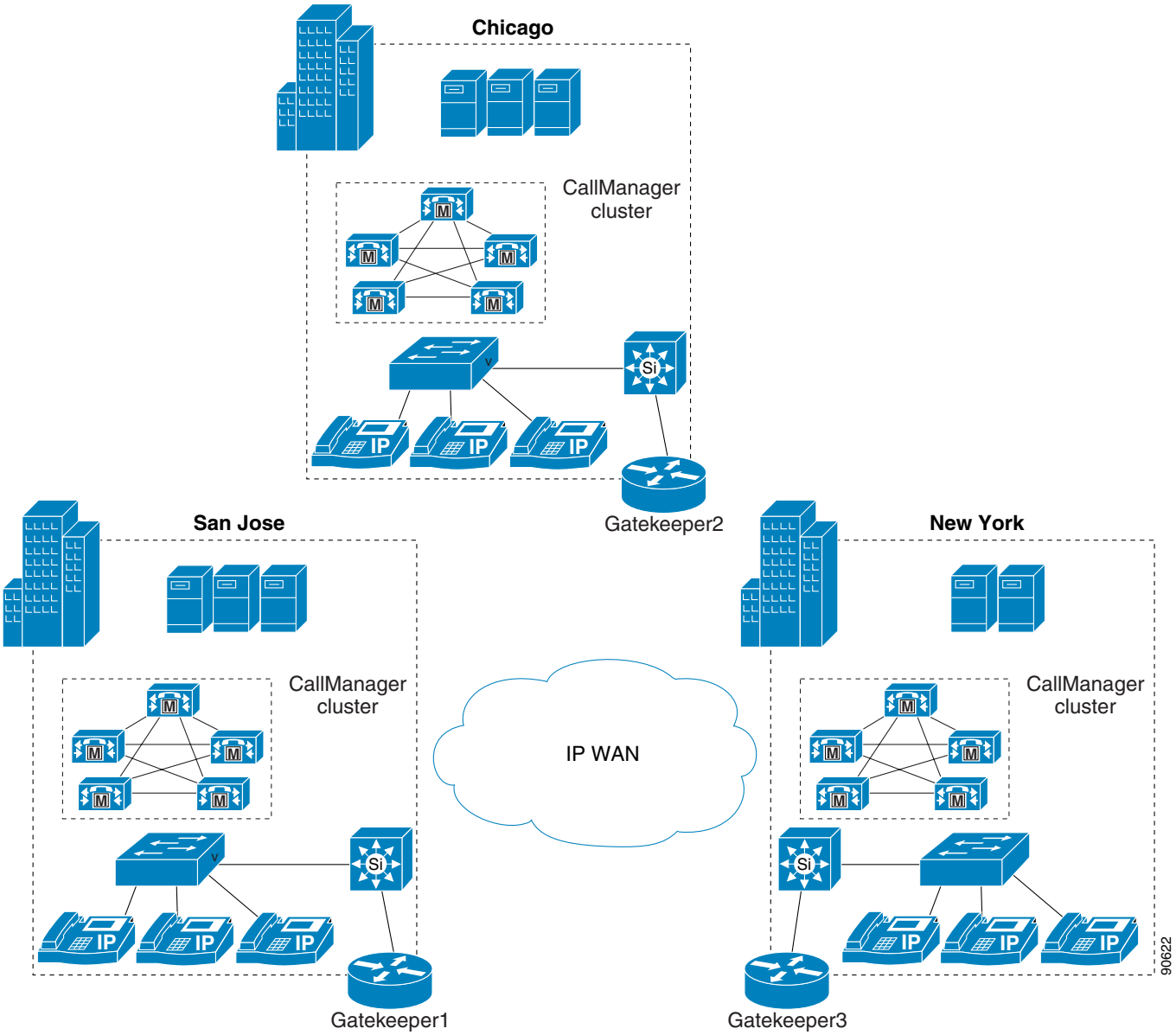
- CPU utilization
- Memory utilization
- Number of active calls
- Number of registered endpoints

With the support of gatekeeper clustering (Alternate-Gatekeeper) and Cisco CallManager Release 3.3, stateful redundancy and load balancing is available. Gatekeeper clustering provides:

- Local and remote clusters
- Up to five gatekeepers in a local cluster
- Gatekeepers in local clusters can be located in different subnets or locations
- No failover delay (Because the alternate gatekeeper is already aware of the endpoint, it does not have to go through the full registration process.)
- Gatekeepers in a cluster pass state information and provide load balancing for:
  - CPU utilization
  - Memory utilization
  - Number of active calls
  - Number of registered endpoints

Figure 6-11 shows three sites with Cisco CallManager distributed call processing and three distributed gatekeepers configured in a local cluster.

Figure 6-11 Gatekeeper Clustering



In Figure 6-11, Gatekeeper 2 is the backup for Gatekeeper 1, Gatekeeper 3 is the backup for Gatekeeper 2, and Gatekeeper 1 is the backup for Gatekeeper3.

Example 6-8 shows the configuration for Gatekeeper 1 (SJC), and Example 6-9 shows the configuration for Gatekeeper 2 (CHC). The configuration for Gatekeeper 3 (NYC) is not shown because it is very similar to the other two.

Example 6-8 Gatekeeper Clustering Configuration for Gatekeeper 1

```
gatekeeper
 zone local SJC_Voice cisco.com 10.1.1.1
 zone local CHC_Voice3 cisco.com
 zone local NYC_Voice2 cisco.com
!
```

```
 zone cluster local SJCVoice_Cluster SJC_Voice
  element SJC_Voice2 10.1.2.1 1719
  element SJC_Voice3 10.1.3.1 1719
 !
 zone cluster local CHCVoice_Cluster CHC_Voice3
  element CHC_Voice2 10.1.3.1 1719
  element CHC_Voice  10.1.2.1 1719
 !
 zone cluster local NYCVoice_Cluster NYC_Voice2
  element NYC_Voice3 10.1.2.1 1719
  element NYC_Voice 10.1.3.1 1719
 !
 zone prefix SJC_Voice 40852.....
 zone prefix NYC_Voice 21251.....
 zone prefix CHC_Voice 72067.....
 gw-type-prefix 1#* default-technology
 load-balance cpu 80 memory 80
 bandwidth interzone SJC_Voice 192
 bandwidth interzone NYC_Voice2 160
 bandwidth interzone CHC_Voice3 160
 arq reject-unknown-prefix
 no shutdown
```

Example 6-9 Gatekeeper Clustering Configuration for Gatekeeper 2

```
gatekeeper
 zone local CHC_Voice cisco.com 10.1.2.1
 zone local SJC_Voice2 cisco.com
 zone local NYC_Voice3 cisco.com
!
 zone cluster local CHCVoice_Cluster CHC_Voice
  element CHC_Voice2 10.1.3.1 1719
  element CHC_Voice3  10.1.1.1 1719
!
 zone cluster local SJCVoice_Cluster SJC_Voice2
  element SJC_Voice 10.1.1.1 1719
  element SJC_Voice3 10.1.3.1 1719
 !
  zone cluster local NYCVoice_Cluster NYC_Voice3
  element NYC_Voice2 10.1.1.1 1719
  element NYC_Voice 10.1.3.1 1719
 !
 zone prefix SJC_Voice 40852.....
 zone prefix NYC_Voice 21251.....
 zone prefix CHC_Voice 72067.....
 gw-type-prefix 1#* default-technology
 load-balance cpu 80 memory 80
 bandwidth interzone CHC_Voice 160
 bandwidth interzone SJC_Voice2 192
 bandwidth interzone NYC_Voice3 160
 arq reject-unknown-prefix
 no shutdown
```

The following notes also apply to Example 6-8 and Example 6-9:

- Each Cisco CallManager cluster has a local zone configured to support Cisco CallManager trunk registrations.
- A cluster is defined for each local zone, with backup zones on the other gatekeepers listed as elements. Elements are listed in the order in which the backups are used.
- A zone prefix is configured for each zone to allow inter-zone and inter-cluster call routing.

- The **gw-type-prefix 1# default-technology** command allows all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.
- The **load-balance cpu 80 memory 80** command limits CPU and memory usage. If the router hits either limit, all new requests are denied and the first backup in the list is used until utilization drops below the threshold.
- Bandwidth statements are configured for each site. Cisco recommends that you use the **bandwidth interzone** command. Using the **bandwidth total** command can cause issues in some configurations.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Cisco CallManager trunks.

All gatekeepers in the cluster display all Cisco CallManager trunk registrations. For trunks that use the gatekeeper as a primary resource, the flag field is blank. For trunks that use another gatekeeper in the cluster as their primary gatekeeper, the flag field is set to A (alternate). Having all endpoints registered as primary or alternate allows all calls to be resolved locally without having to send a location request (LRQ) to another gatekeeper.

Example 6-10 shows the output from the **show gatekeeper endpoints** command on Gatekeeper 1 (SJC).

Example 6-10 Output for Gatekeeper Endpoints

```
                    GATEKEEPER ENDPOINT REGISTRATION
                    ================================
CallSignalAddr  Port  RASSignalAddr   Port  Zone Name         Type        Flags
--------------- ----- --------------- ----- ---------         ----        -----
10.1.1.12       1307  10.1.1.12       1254  SJC_Voice         VOIP-GW
    H323-ID: SJC-to-GK-trunk_1
10.1.1.12       4422  10.1.2.12       4330  SJC_Voice         VOIP-GW
    H323-ID: SJC-to-GK-trunk_2
10.1.2.12       4587  10.1.2.12       4330  NYC_Voice3        VOIP-GW     A
    H323-ID: CHC-to-GK-trunk_1
10.1.3.21       2249  10.1.3.21       1245  CHC_Voice2        VOIP-GW     A
    H323-ID:  NYC-to-GK-trunk_1
Total number of active registrations = 4
```

Directory Gatekeeper Redundancy

You can implement directory gatekeeper redundancy by using HSRP or by configuring two identical directory gatekeepers. When a gatekeeper is configured with multiple remote zones using the same zone prefix, the gatekeeper can use either of the following methods:

- Sequential LRQs (default)

  Redundant remote zones (matching zone prefixes) are assigned a cost, and LRQs are sent to the matching zones in order based on the cost values. Using sequential LRQs saves WAN bandwidth by not blasting LRQs to all matching gatekeepers.

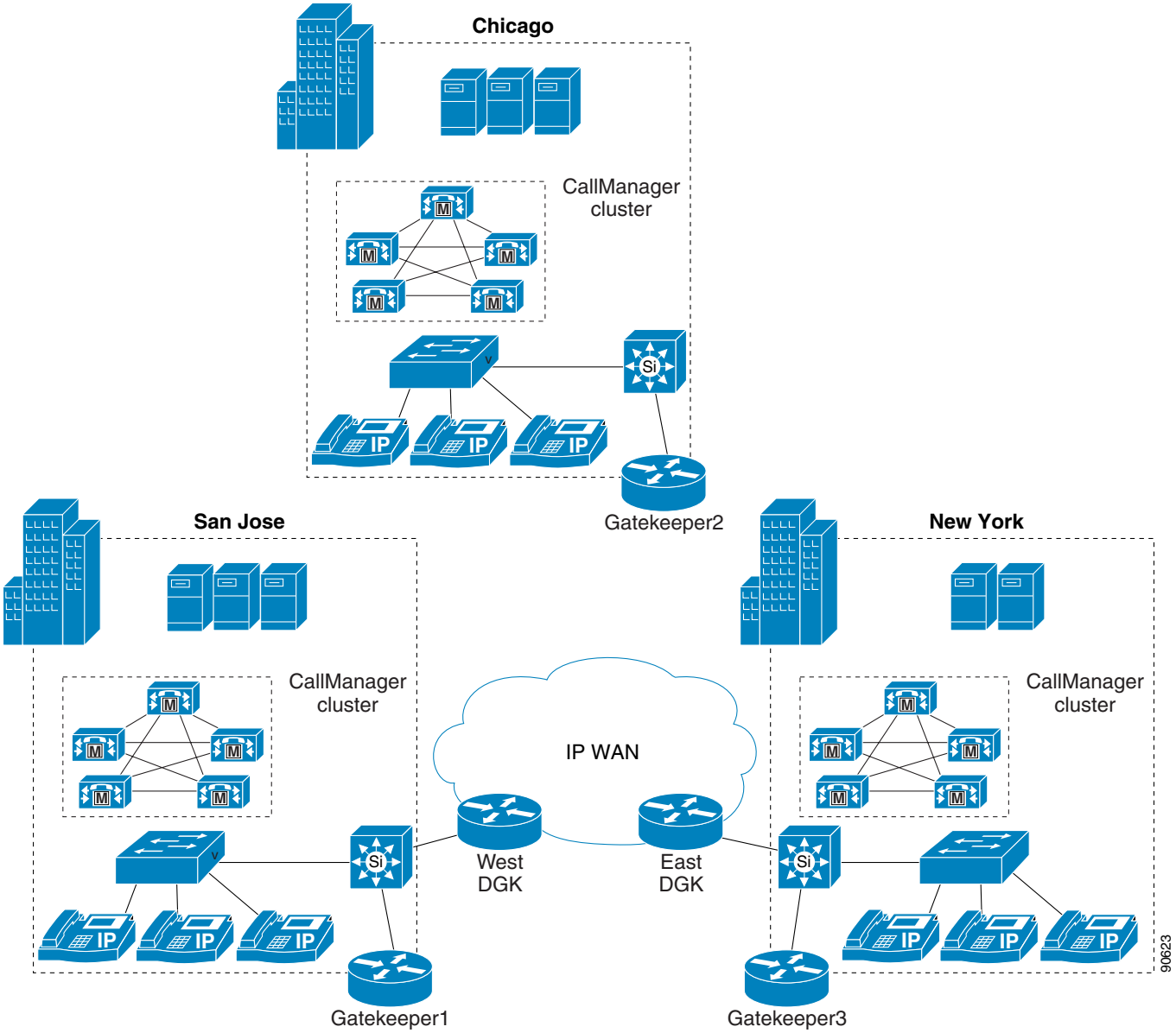
- LRQ Blast

  LRQs are sent to redundant zones (matching zone prefixes) simultaneously. The first gatekeeper to respond with an Location Confirm (LCF) is the one that is used.

Cisco recommends that you use two active directory gatekeepers with sequential LRQs, thus allowing directory gatekeepers to be placed in different locations. Using HSRP requires both directory gatekeepers to be located in the same subnet, and only one gatekeeper can be active at any time.

Figure 6-12 illustrates a Cisco CallManager distributed call processing environment with two active directory gatekeepers.

Figure 6-12 Redundant Directory Gatekeepers



Example 6-11 and Example 6-12 show the configurations for the two directory gatekeepers in Figure 6-12.

Example 6-11 Configuration for West Directory Gatekeeper

```
gatekeeper
 zone local DGKW customer.com 10.1.10.101
 zone remote GK-SJC customer.com 10.1.10.100
 zone remote GK-CHC customer.com 10.1.11.100
 zone remote GK-NYC customer.com 10.1.12.100
 zone prefix GK-SJC 408*
```

```
 zone prefix GK-CHC 720*
 zone prefix GK-NYC 212*
 lrq forward-queries
 no shutdown
```

Example 6-12 Configuration for East Directory Gatekeeper

```
gatekeeper
 zone local DGKE customer.com 10.1.12.101
 zone remote GK-SJC customer.com 10.1.10.100
 zone remote GK-CHC customer.com 10.1.11.100
 zone remote GK-NYC customer.com 10.1.12.100
 zone prefix GK-SJC 408*
 zone prefix GK-CHC 720*
 zone prefix GK-NYC 212*
 lrq forward-queries
 no shutdown
```

The following notes also apply to Example 6-11 and Example 6-12:

- Both directory gatekeepers are configured exactly the same.
- A local zone is configured for the directory gatekeeper.
- Remote zones are configured for each remote gatekeeper.
- Zone prefixes are configured for both remote zones for inter-zone call routing. The wildcard (*) is used in the zone prefix to simplify configuration. Calls are not routed to the DGK zone, so a prefix is not required for it.
- The **lrq forward-queries** command allows the directory gatekeeper to forward an LRQ received from another gatekeeper.

Note Directory gatekeepers do not contain any active endpoint registrations and do not supply any bandwidth management.

Example 6-13, Example 6-14, and Example 6-15 show the configurations for Gatekeepers 1 to 3 in Figure 6-12.

Example 6-13 Configuration for Gatekeeper 1 (SJC)

```
gatekeeper
 zone local GK-SJC customer.com 10.1.10.100
 zone remote DGKWcustomer.com 10.1.10.101
 zone remote DGKE customer.com 10.1.12.101
 zone prefix GK-SJC 408.......
 zone prefix DGKW ..........
 zone prefix DGKE ..........
 bandwidth remote 192
 gw-type-prefix 1# default-technology
 arq reject-unknown-prefix
 no shutdown
```

Example 6-14 Configuration for Gatekeeper 2 (CHC)

```
gatekeeper
 zone local GK-CHC customer.com 10.1.11.100
 zone remote DGKE customer.com 10.1.12.100
 zone remote DGKW customer.com 10.1.10.100
 zone prefix GK-CHC 720.......
 zone prefix DGKE ..........
 zone prefix DGKW ..........
 bandwidth remote 160
 gw-type-prefix 1# default-technology
 arq reject-unknown-prefix
 no shutdown
```

Example 6-15 Configuration for Gatekeeper 3 (NYC)

```
gatekeeper
 zone local GK-NYC customer.com 10.1.12.100
 zone remote DGKE customer.com 10.1.12.100
 zone remote DGKW customer.com 10.1.10.100
 zone prefix GK-NYC 212.......
 zone prefix DGKE ..........
 zone prefix DGKW ..........
 bandwidth remote 160
 gw-type-prefix 1# default-technology
 arq reject-unknown-prefix
 no shutdown
```

The following notes also apply to Example 6-13, Example 6-14, and Example 6-15:

- Each Cisco CallManager cluster has a local zone configured to support Cisco CallManager trunk registrations.
- Remote zones are configured for each directory gatekeeper.
- Zone prefixes are configured for the local zone and both remote zones for inter-zone call routing. Both directory gatekeeper prefixes are 10 dots. Sequential LRQs are used by default when matching zone prefixes are configured. The gatekeeper sends an LRQ to the directory gatekeeper with the lowest cost; if there is no response, the gatekeeper tries the second directory gatekeeper.
- The **bandwidth remote** command is used to limit bandwidth between the local zone and any other remote zone.
- The **gw-type-prefix 1# default-technology** command allows all locally unresolved calls to be forwarded to a device registered with a technology prefix of 1# in the local zone. In this example, all Cisco CallManager trunks have been configured to register with a 1# prefix.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Cisco CallManager trunks.

Dial Plan

This chapter summarizes dial plan design considerations and guidelines for the following deployment models:

Dial Plan Guidelines for All Deployment Models

The following dial plan guidelines apply to all deployment models:

Also refer to the section on Dial Plan Weights, page 6-5, to ensure that your dial plan remains within acceptable design guidelines.

External Route Configuration

Cisco CallManager automatically routes calls to destinations within the same cluster. For external destinations such as PSTN gateways, H.323 gatekeepers, or other Cisco CallManager clusters, use the following elements in Cisco CallManager to configure explicit routing:

Route Patterns

Route Patterns are strings of digits and wildcards, such as 9.[2-9]XXXXXX, configured in Cisco CallManager to route calls to external entities.

The @ Wildcard

- The @ wildcard is a special macro function that expands into a series of patterns representing the entire North American Numbering Plan. For example, configuring a single unfiltered route pattern such as 9.@ really adds 166 individual route patterns to the Cisco CallManager dial plan configuration.
- For large centralized call processing deployments with distributed PSTN gateways, configure explicit route patterns for local PSTN access instead of using the @ wildcard. This configuration will reduce the size of the resulting dial plan database.

Route Filters

- Use route filters only with the @ route pattern (most commonly, with 9.@) to reduce the number of route patterns created by the @ wildcard.
- The logical expression you enter with the route filter can be up to 1024 characters, excluding the NOT-SELECTED fields.
- As you increase the number of logical clauses in a route filter, the refresh time of the configuration page also increases and can become unacceptably long.
- For large-scale deployments, use explicit route patterns rather than the @ wildcard and route filters. This practice also facilitates management and troubleshooting because all patterns configured in Cisco CallManager are easily visible from the Route Pattern configuration page.

International and Variable-Length Route Patterns

- International destinations are usually configured using the ! wildcard, which represents any quantity of digits. For example, in North America the route pattern 9.011! is typically configured for international calls.
- The ! wildcard is also used for deployments in countries where the dialed numbers can be of varying lengths (for example, in Germany). In such cases, Cisco CallManager does not know when the dialing is complete and will wait for 15 seconds before sending the call. You can reduce this delay in any of the following ways:
  - Reduce the T302 timer (Service Parameter TimerT302_msec) to indicate end of dialing, but do not set it lower than 4 seconds to prevent premature transmission of the call before the user is finished dialing.
  - Configure a second route pattern followed by the # wildcard (for example, 9.011!# for North America), and instruct the users to dial # to indicate end of dialing. This action is analogous to hitting the “send” button on a cell phone.

Digit Manipulation in Route Patterns

- Configure digit manipulation only in the route group and not in the route pattern.
- Digit manipulation in the route group completely overrides any digit manipulation done in the route pattern.
- If you configure digit manipulation in the route pattern, the Call Detail Record (CDR) records the dialed number after the digit manipulation has occurred. If you configure digit manipulation only in the route group, the CDR records the actual dialed number prior to the digit manipulation.

Calling Line ID

- The calling line ID presentation can be enabled or disable on the gateway and also can be manipulated in the route pattern, based on site requirements.
- If you select the Use Calling Party's External Phone Number Mask option, then the external call uses the calling line ID specified for the IP phone placing the call. If you do not select this option, then the mask placed in the Calling Party Transform Mask field is used to generate the calling party ID.

Route Lists

A route list is a prioritized list of eligible paths (route groups) for an outbound call. Typically, a route list is associated with a remote location, and multiple route patterns may point to it. A typical use of a route list is to specify two paths for a remote destination, where the first-choice path is across the IP WAN and the second-choice path is through the local PSTN gateways.

Route lists have the following characteristics:

- Multiple route patterns may point to the same route list.
- A route list is a prioritized list of route groups that function as alternate paths to a given destination. For example, you can use a route list to provide least-cost routing, where the primary route group in the list offers a lower cost per call and the secondary route group is used only if the primary is unavailable due to an "all trunks busy" condition or insufficient IP WAN resources.
- Each route group in the route list can have its own digit manipulation. For example, if the route pattern is 9.@ and a user dials 9-1-408-555-4000, the IP WAN route group can strip off the 9-1 while the PSTN route group may strip off just the 9.
- Multiple route lists can contain the same route group. The digit manipulation for the route group is associated with the specific route list that “points” to the route group.
- If you are performing several digit manipulations in a route pattern or a route group, the order in which the transformations are performed can impact the resulting E.164 address.
  Cisco CallManager performs the following major types of digit manipulations in the order indicated:

1. Discarding digits
2. Called party transformations
3. Prefixing digits

Route Groups

Route groups control and point to specific devices, which are typically gateways (MGCP or H.323) or H.323 trunks to a gatekeeper or a remote Cisco CallManager cluster. (In Cisco CallManager Release 3.2 and earlier, the role of the H.323 trunk was performed by the "Anonymous Device" gateway and by H.323 gateways configured using the Intercluster Trunk protocol.)

You can assign an order to the devices within a route group, and Cisco CallManager will send calls to the devices in that specified order. If you want to use round-robin ordering for outgoing calls, set the orders of all devices in a route group to 1, and set the service parameter ReorderRouteList to TRUE. With these settings, all route group members with the same order priority will then take turns being selected to route calls.

Route Group Devices

The route group devices are the endpoints accessed by route groups, and they typically consist of gateways or H.323 trunks to a gatekeeper or to remote Cisco CallManagers. You can configure the following types of devices in Cisco CallManager:

- MGCP gateways
- H.323 gateways
- H.225 trunk, gatekeeper controlled — trunk to standard H.323 gateways, via a gatekeeper
- Intercluster trunk, not gatekeeper controlled — direct trunk to another Cisco CallManager cluster
- Intercluster trunk, gatekeeper controlled — trunk to other Cisco CallManager clusters and/or H.323 gateways, via a gatekeeper

Note Both the H.225 and intercluster trunk (gatekeeper controlled) will automatically discover if the other endpoint is a standard H.323 gateway or a Cisco CallManager and will select H.225 or Intercluster Trunk protocol accordingly. (This auto-discovery mechanism also applies to "Anonymous Device" gateways configured with intercluster protocol for Cisco CallManager Release 3.2.) If you are setting up a direct trunk to a cluster prior to Cisco CallManager Release 3.1, then choose Intercluster Trunk protocol.

Calling Restrictions

To implement calling restrictions, configure the following elements in Cisco CallManager:

- Calling Search Spaces, page 7-4
- Partitions, page 7-5

A *partition* is a group of devices with similar accessibility, and a *calling search space* defines which partitions are accessible to a particular device. A device can call only those devices located in the partitions that are part of its calling search space.

Calling Search Spaces

A calling search space defines which partitions are accessible to a particular device. Devices that are assigned a certain calling search space can access only the partitions listed in that calling search space. Attempts to dial a DN in a partition outside that calling search space will fail, and the caller will hear a busy signal.

If you configure a calling search space both on an IP phone line and on the device (phone) itself, Cisco CallManager concatenates the two calling search spaces and places the line's calling search space in front of the device's calling search space. If the same route pattern appears in two partitions, one contained in the line's calling search space and one contained in the device's calling search space, then Cisco CallManager selects the route pattern listed first in the concatenated list of partitions (in this case, the route pattern associated with the line's calling search space).

Note Prior to Cisco CallManager Release .3.1, the concatenation was performed in the reverse order, and the device's calling search space came first, followed by the line's calling search space.

Note Cisco strongly recommends avoiding the configuration of equally matching patterns in partitions that are part of the same calling search space or part of different calling search spaces configured on the same phone. This practice avoids the difficulties related to predicting dial plan routing when the calling search space's partition order is used as a tie breaker.

The maximum length of the combined calling search space (device plus line) is 1024 characters, including separator characters between each partition name. (For example, the string "partition_1:partition_2:partition_3" contains 35 characters.) Thus, the maximum number of partitions in a calling search space varies, depending on the length of the partition names. Also, because the calling search space clause combines the calling search space of the device and that of the line, the maximum character limit for an individual calling search space is 512 (half of the combined calling search space clause limit of 1024 characters).

Therefore, when you are creating partitions and calling search spaces, keep the names of partitions short relative to the number of partitions that you plan to include in a calling search space. For more details on configuring calling search spaces, refer to the *Cisco CallManager Administration Guide*, available online at

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/3_3/sys_ad/3_3_2/ccmcfg/index.htm

Before you configure any partitions or calling search spaces, all DNs reside in a special partition named <None>, and all devices are assigned a calling search space also named <None>. When you create custom partitions and calling search spaces, any calling search space you create also contains the <None> partition, while the <None> calling search space contains **only** the <None> partition.

Note Any dial plan entry left in the <None> partition is implicitly reachable by any device making a call. Therefore, to avoid unexpected results, Cisco strongly recommends that you do not leave dial plan entries in the <None> partition.

Cisco recommends the following best practices for configuring calling search spaces on IP phones:

- To ensure that dialing privileges are uniform for all lines on a given phone, you may configure the calling search space on the IP phone itself and not on the individual lines of the phone. This practice prevents users from selecting another line on the phone to bypass calling restrictions.
- When configuring call forward features on an IP phone line, do not select a calling search space that can reach the PSTN. This practice prevents users from forwarding their IP phone lines to a long-distance number and dialing their local IP phone number to bypass long-distance toll charges on personal calls.

Partitions

The dial plan entries that you may place in a partition include IP phone directory numbers, translation patterns, route patterns, CTI route points, and voice mail ports.

If two or more dial plan entries (directory numbers, route patterns, or so forth) overlap, Cisco CallManager selects the entry with the closest match (most specific match) to the dialed number. For example, route pattern 11XX is a closer match for the dialed digits 1111 than is the translation pattern 1XXX, so Cisco CallManager would use the route pattern 11XX to route the call in this case.

In cases where two dial plan entries match the dialed pattern equally, the behavior depends on whether the matching entries are in the same or different partitions, as follows:

- Multiple equal-precision matches in different partitions:

  In this case, Cisco CallManager selects the dial plan entry that appears first in the calling search space of the device making the call. For example, assume that route pattern 11XX is part of Partition_A, translation pattern 11XX is part of Partition_B, and the calling search space of the calling device lists the partitions in the order Partition_B:Partition_A. In this example, Cisco CallManager would select translation pattern 11XX as the matching entry because its partition (Partition_B) is listed first in the calling search space of the calling device.

- Multiple equal-precision matches in the same partition:

  In this case, Cisco CallManager selects the entry that is listed first in its local dial plan database.

Note You cannot configure the order in which the dial plan database lists dial plan entries. Therefore, Cisco strongly recommends that you avoid any possibility of equal-precision matches co-existing within the same partition because the resulting dial plan logic is not predictable in such cases.

Building Classes of Service

You can define classes of service for different telephony users by combining partitions and calling search spaces with external route patterns, as follows:

- Place external route patterns in partitions associated with the destinations they can call. While you could place all route patterns in a single partition, you can achieve more refined call restriction policies by associating the route patterns with partitions according to the destinations they can call. For example, if you place local and international route patterns in the same partition, then all users can reach both local and international destinations, which might not be desirable.
- Configure each calling search space to be able to reach only the partitions associated with its call restriction policy. For example, configure the local calling search space to point to the internal and local partitions, so that users assigned to this calling search space can place only local calls.

Translation Patterns

Translation patterns follow the same general rules and use the same wildcards as route patterns. (See Route Patterns, page 7-2.) As with route patterns, you assign a translation pattern to a partition. However, when the dialed digits match the translation pattern, Cisco CallManager does not route the call to an outside entity such as a gateway; instead, it performs the translation then routes the call again, this time using the calling search space configured within the translation pattern.

Translation patterns can be used to provide inter-site dialing in the presence of overlapping extensions. For instance, if both Site 1 and Site 2 have extensions in the range 1XXX, partitions must be used to separate their overlapping directory numbers. To allow communication between sites, a set of translation patterns (one per site) is defined in a common partition that is visible to all users. When the user with extension 1000 at Site 1 wishes to dial the user with extension 1000 at Site 2, the user at Site 1 first dials the inter-site access code (for example, 8) followed by the destination site code (for example, 2) followed by the 4-digit extension of the other party (1000 in this case). This string, 821000, matches a configured translation pattern that strips 82 and delivers 1000 to the Site2 internal calling search space, which has access to Site 2's directory numbers.

For more details on how to design dial plans in the presence of overlapping extensions, see Centralized Call Processing with Overlapping Extensions, page 7-12.

Dial Plan Guidelines for Single-Site Deployments

A single- site deployment is the simplest with respect to the dial plan because it typically uses only one path, the PSTN, for all external calls. Although each dial plan has it own special characteristics, the following general considerations apply to single-site deployments:

- You can use a single route list that contains only one route group because there is only one path to the PSTN. You could configure multiple route groups if some PSTN trunks connect to Carrier A and some to Carrier B, where Carrier A might be the preferred carrier.
- You can use a single PSTN gateway. To add capacity or provide redundancy, you could configure multiple gateways as devices within the route group.
- You can configure the route pattern 9.[2-9]XXXXXX to allow 7-digit dialing and to implement a call restriction policy that limits some users to dial local calls only. Place the 9.[2-9]XXXXXX route pattern in a different partition than the 9.1[2-9]XX[2-9]XXXXXX route pattern, and configure the restricted users' calling search space to contain only the partition with the 7-digit dialing route pattern. An alternative to configuring the 9.[2-9]XXXXXX route pattern would be to apply the pre-configured “7-digit dialing” route filter to the 9.@ route pattern.
- Place the PreDot and Trailing # digit discard instructions in the route group. The PreDot instruction strips the access code 9, and the Trailing # instruction removes the # that users may dial to signify end of dialing for international calls.
- Place route patterns in partitions based on the granularity of the policies created by the calling search spaces. This is why the 9.[2-9]XXXXXX route pattern is placed in its own partition and not with the national and international route patterns.
- Configure a separate calling search space for each call restriction policy at your site, and point the calling search space to the respective partitions as needed.
- Place all IP phones in an internal partition that is reachable from all calling search spaces This configuration allows any IP phone to dial any other IP phone.

Dial Plan Guidelines for Multi-Site IP WAN Deployments with Centralized Call Processing

In a multi-site IP WAN deployment with centralized call processing, the call processing and applications (such as voice mail) are centralized but the dial plan is designed for each site individually. Typically, PSTN gateways are distributed across the remote sites, and users at those sites expect their calls to go out the local PSTN gateway when they dial the PSTN access code. In addition, emergency calls such as 911 must always go through the local PSTN gateway.

Note This section assumes that there are no overlapping extensions among the sites. For design considerations in the presence of overlapping extensions, see Centralized Call Processing with Overlapping Extensions, page 7-12.

Route Pattern Structure

In most cases, each site requires that its emergency calls use the local PSTN. To implement this requirement, configure a separate set of route patterns for each remote site, and place the route patterns in a partition that can be reached only by the users at that particular remote site. This recommendation assumes that the PSTN gateways are located at each site. If all PSTN gateways were centralized at the hub site, the dial plan configuration would become identical to that for a single-site deployment.

Partitions and Calling Search Spaces

In most cases, users in a centralized call processing deployment expect to be able to call a remote site by simply dialing the internal extension of a remote user. Use the following guidelines to implement this type of dial plan:

- To facilitate on-net dialing between sites, place all IP phones in an on-cluster partition that can be reached from the calling search spaces of all remote sites. Note that this model does not allow for overlapping dial extensions among remote sites.
- Give each remote site its own set of partitions and route patterns. The number of partitions per remote site depends on the number of calling restriction policies associated with the route patterns.
- Give each site its own set of calling search spaces for its IP phones. The calling search spaces point to the on-cluster partition as well as to the appropriate local route pattern partitions.
- Use the following formulas to calculate the total number of calling search spaces and partitions needed:

  Total Partitions = (Number of calling restriction policies) * (Number of sites) + (1 Partition for all IP phones)

  Total Calling Search Spaces = (Number of calling restriction policies per site) * (Number of sites)

An Alternative Approach to Configuring Calling Search Spaces

The approach outlined in the preceding section implements dialing privileges through the use of multiple branch-specific calling search spaces (one for each branch's call restriction policy). This configuration is required because each branch needs to use different gateways for local and emergency calls.

An alternative approach is to create a single branch-specific calling search space and assign it to all the branch phones through the use of the Device Calling Search Space.

Use the following general approach to configure calling restrictions with a single branch-specific calling search space:

- Create an unrestricted calling search space for each location and assign it to the phone's device calling search space. This calling search space should contain partitions featuring route patterns that route the calls to the appropriate gateway for the phone's location (for example, a co-located branch gateway for emergency services and a centralized gateway for long distance calls).
- Create calling search spaces containing partitions featuring blocked route patterns for those types of calls not part of the user's dialing privileges, and assign them to the user's lines. For instance, if a user has access to all types of calls except international, that user's line (or lines) should be configured with a calling search space featuring a blocked route pattern for 9.011!

This approach has the significant advantage that only a single, un-restricted calling search space is required for each location (that is, one per branch). Because the dialing privileges are implemented through the use of blocked route patterns (which are not locally significant), the same set of blocking calling search spaces can be used in all branches.

For a system with N branches and, for example, 5 different classes of service, N+5 calling search spaces are required. This number is in contrast with the 5∗N calling search spaces required by the typical approach.

Special Considerations for Extension Mobility

When using the Extension Mobility feature, the dialing restrictions of a phone are a function of the logged-in (or logged-out) status of the phone. Typically, a logged-out phone should be restricted to calling other phones and services (such as 911), but access to local or toll calls through the PSTN are restricted. Conversely, a phone where a user is logged-in should allow calls according to that user's dialing privileges and should route those calls to the appropriate gateway (for example, a co-located branch gateway for local calls).

Use the following general approach to configure calling restrictions when using Extension Mobility:

- Create an unrestricted calling search space for each location and assign it to the phone's device calling search space. This calling search space should contain partitions featuring route patterns that route the calls to the appropriate gateway for the phone's location (for example, a co-located branch gateway for emergency services and a centralized gateway for long distance calls).
- Create a calling search space with partitions featuring blocked route patterns and assign it to the lines in the Default Logout Device Profile. The blocked route patterns should block all calls except those to be allowed when no user is logged in (for example, emergency services and internal extensions).
- Create user device profiles with line calling search spaces containing partitions featuring blocked route patterns for those types of calls that are not part of the user's dialing privileges. For instance, if a user has access to all types of calls except international, that user device profile should be configured with a calling search space featuring a blocked route pattern for 9.011!

Automated Alternate Routing

The automated alternate routing (AAR) feature enables Cisco CallManager to establish an alternate path for the voice media when the preferred path between two intra-cluster endpoints runs out of available bandwidth, as determined by the locations mechanism for call admission control (CAC).

The AAR feature applies primarily to centralized call processing deployments. For instance, if a phone in branch A calls a phone in branch B and the available bandwidth for the WAN link between the branches is insufficient (as computed by the locations mechanism), AAR can reroute the call through the PSTN. The audio path of the call would be IP-based from the calling phone to its local (branch A) PSTN gateway, TDM-based from that gateway through the PSTN to the branch B gateway, and IP-based from the branch B gateway to the destination IP phone.

AAR can be transparent to the users. You can configure AAR so that users dial only the on-net (for example, 4-digit) directory number of the called phone and no additional user input is required to reach the destination through the alternate network (such as the PSTN).

Note AAR does not support CTI route points as the origin or the destination of calls. Also, AAR does not support the Extension Mobility feature.

You must provide the following main elements for AAR to function properly:

- Establish the PSTN Number of the Destination, page 7-10
- Prefix the Required Access Codes, page 7-10
- Select the Proper Dial Plan and Route, page 7-10

Establish the PSTN Number of the Destination

The rerouting of calls requires using a destination directory number (DN) that is routable through the alternate network (for example, the PSTN). AAR uses the dialed digits to establish the on-cluster destination of the call and then combines them with the called party's External Phone Number Mask. The combination of these two elements must yield a fully qualified number that is routable by the alternate network.

For example, assume phone A in San Francisco (DN = 2345) dials an on-net DN (1234) configured on phone B located in New York. If locations-based call admission control denies the call, AAR retrieves the External Phone Number Mask of the New York phone (212-555-XXXX) and uses it to derive the fully qualified number (212-555-1234) that is routable on the PSTN.

The PSTN routing of a call from San Francisco to New York requires a "1" as a prefix to the phone number. Cisco recommends that you do *not* include this prefix as part of the External Phone Number Mask of phones because it would be displayed as part of the Calling Party Identification (CallerID) for any calls made by the phones to an off-net destination. Instead, Cisco recommends that you add the "1" as part of the AAR group configuration.

Prefix the Required Access Codes

The destination number might require a prefix for an off-net access code (for example, 9) to be routed properly by the origination branch's dial plan. Furthermore, if the point of origin is located in a different Numbering Plan Area (NPA, or area code), a prefix of "1" might be required as part of the dialed string.

When configuring AAR, you place the DNs in AAR groups. For each pair of AAR groups, you can then configure prefix digits to add to the DNs for calls between the two groups, including prefix digits for calls originating and terminating within the same AAR group.

As a general rule, place DNs in the same AAR group if they share all of the following characteristics:

- A common off-net access code (for example, 9)
- A common PSTN dialing structure for inter-area calls (for example, 1-NPA-NXX-XXXX in North America)
- A common External Phone Number Mask format

For example, assume that both the San Francisco and New York sites share all of the preceding characteristics. We could place the DNs for San Francisco and New York into a single AAR group and configure the group such that AAR calls placed within this same AAR group are prefixed with 91. For phone A in San Francisco to reach phone B in New York (at 212-555-1212), the AAR group configuration prefixes 91 to the dialed string, yielding a completed string of 91-212-555-1212.

Select the Proper Dial Plan and Route

AAR calls should egress through a gateway within the same location as the calling phone, thus causing to the completed dial string be sent through the origination site's dial plan. To ensure that this is the case, select the appropriate AAR calling search space on the device configuration page in Cisco CallManager

Administration. Configure the off-net dial plan entries (for example, route patterns) in the AAR calling search space to point to co-located gateways and to remove the access code before presenting the call to the PSTN.

For example, phones at the San Francisco site can be configured with an AAR calling search space that permits long distance calls dialed as 91-NPA-NXX-XXXX but that delivers them to the San Francisco gateway with the access code (9) stripped.

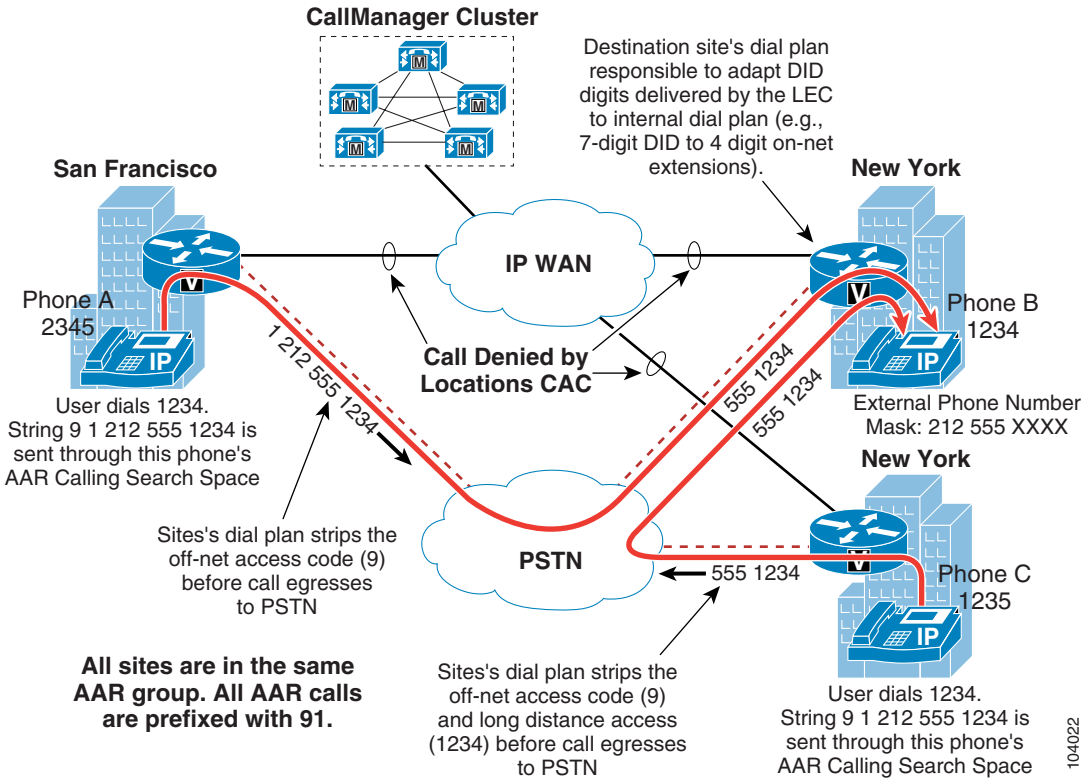
Special Considerations for Sites Located Within the Same Local Dialing Area

In some instances, the AAR dial string might have to be modified locally to allow for local area dialing.

For example, assume two separate sites located in New York share the same NPA of 212. (See Figure 7-1.) In this case, a number dialed as 91-212-555-1234 would have to be transformed to 9-555-1234.

This transformation is best done with a site-specific translation pattern of 91212.555XXXX (to strip the pre-dot digits and prepend 9). This translation pattern should be placed in a member partition of the AAR calling search space for the New York sites only; the San Francisco site still needs to reach this same destination as 91-212-555-1234. This translation pattern should also be placed in the New York sites' dial plan to provide proper routing of locally reachable numbers dialed as long distance calls. The New York sites' dial plan is responsible for accepting 9-555-1234 as a valid string and transforming it to 555-1234 before delivering the call to the PSTN.

Figure 7-1 Dialed Number Transformations for AAR Calls Between Sites



Note The AAR functionality is not triggered upon detection that the destination phone is unreachable. Therefore, WAN failures do not trigger AAR functionality.

Centralized Call Processing with Overlapping Extensions

A multi-site WAN deployment with overlapping extensions is a special case of the centralized call processing model that uses the same IP phone extensions at several different sites. The following considerations apply to this deployment model:

- Several sites can share a single Cisco CallManager cluster and a single voice mail or unified messaging system (similarly to the way a simple centralized call processing deployment shares resources).
- Intra-site calls can be placed using abbreviated dialing (typically 4 or 5 digits).
- Inter-site calls can be placed using either full E.164 dialing or an access code followed by a site code and the extension. For example, if the internal site-to-site access code is 8 and two digits are used for the site code, a user can dial 8-55-20000 to call extension 20000 in site 55.

Support for overlapping extensions increases the complexity of the Cisco CallManager dial plan configuration. However, this type of dial plan may be required in scenarios where extension DNs cannot be changed but abbreviated dialing must be preserved (such as in pre-existing legacy systems, company acquisitions, or mergers).

Note From a dial plan perspective, the same considerations made for enterprise deployments with overlapping extensions also apply to single-site multi-tenant deployments.

Partitions and Calling Search Spaces

In an overlapping dial plan, place the IP phones at each site in a separate partition. For each site, you can define additional partitions to hold the local PSTN route patterns. The number of these partitions you need depends on the number of classes of service (or calling policies) you want to implement, but they generally are one of the following types:

- Site-specific partitions
  - Hold the DNs of all IP phones located at the site
  - Hold the local route patterns for emergency calls, local calls, national calls, and international calls
- Common partitions
  - Provide access to shared resources such as voice mail, media resources, and applications
  - Hold the translation patterns needed to provide inter-site calls

You can then assign IP phones to a calling search space that contains a subset of these partitions, according to the calling policy assigned to them.

Note This configuration differs from that adopted for non-overlapping centralized call processing deployments, where all IP phones were placed in the same partition. In this case, the IP phones must be placed in different partitions because their DNs are not unique. If they were all placed in the same partition, they would automatically become shared line appearances.

Outbound Calls

Each site typically requires that emergency and local PSTN calls use the local branch PSTN gateway. You can implement this policy by placing the corresponding route patterns in partitions that are reachable only from that site's IP phones.

Inter-Site Calls

With overlapping dial plans, inter-site calls are placed either by dialing the full E.164 number of the destination or by using an inter-site access code followed by a site code and the extension number. The same design considerations apply in either case.

To provide connectivity between sites and partitions, use translation patterns according to the following guidelines:

- Define one translation pattern for each site, and place them all in the on-cluster partition.
- Each pattern should match a site's E.164 address range.
- The resulting called numbers after translation should match the site's extensions.
- The calling search space where the call is sent after translation should include the partition that contains the site's IP phones.

Incoming Calls

To dispatch incoming PSTN calls to the appropriate extension, you can re-use the translation patterns in the on-cluster partition mentioned previously. It is sufficient to assign all PSTN gateways a calling search space that contains only the on-cluster partition, making sure that the gateways also prepend a 9 to the dialed number to match the already defined translation patterns.

Voice Mail Considerations

Voice mail integration requires special attention to the following requirements in the presence of overlapping extensions:

- Voice mail boxes must have unique identifiers. This means that the IP phone extension cannot be used as a voice mail box, and some sort of digit manipulation is needed to obtain a unique number.
- Message waiting indicator (MWI) messages from the voice mail system must be able to reach the right IP phone even in the presence of non-unique extensions.

The first issue is addressed in Cisco CallManager by the use of the Voice Mail Box Mask field in the Voice Mail Profile Configuration page. This parameter, when configured, is used to communicate with the voice mail system and uniquely identify the user. For example, the Voice Mail Box Mask parameter can be set to the full E.164 number associated with the user.

The second issue is addressed by re-using the translation patterns in the on-cluster partition. If the voice mail system has been configured with the full E.164 numbers, it is sufficient to prepend 9 to the E.164 number in order to match the translation patterns previously defined and to ensure proper inter-site communication. In this way, the MWI messages coming from the voice mail system with the full E.164 number will be translated to the appropriate extension in the specific partition.

Note This scenario requires the configuration of two service parameters within Cisco CallManager. The MultiTenantMwiMode parameter within the Cisco CallManager service must be set to True, and the ValidateDNs parameter within the Cisco Messaging Interface (CMI) service must be set to False.

Dial Plan Guidelines for Multi-Site IP WAN Deployments with Distributed Call Processing

In a multi-site IP WAN deployment with distributed call processing, the dial plan is typically configured to use the IP WAN as the first choice for on-net calls and the PSTN as the second choice if the IP WAN is not available or cannot handle additional voice calls. It is a common requirement for this type of deployment that intra-site calls use some sort of abbreviated dialing (for example, 5-digit dialing), while inter-site calls, either across the IP WAN or to the PSTN, use the full E.164 numbers.

Route Pattern Structure

Although there are many possible route pattern configurations for the distributed call processing model, it is typical to use the same route list to make the path selection for all route patterns that route on-net calls to the IP WAN as the primary voice path.

Observe the following best practices when designing a dial plan for a multi-site WAN with distributed call processing:

- Use the IP WAN for on-net internal company calls only. It is possible to provide remote hop-off to save long distance costs, but this practice complicates the dial plan configuration.
- It is common practice to send the full E.164 address to the gatekeeper and the remote Cisco CallManager or gateway, leaving it up to the terminating device to strip off all but the significant digits. This practice eliminates the need to configure the local (calling) Cisco CallManager with dial length information for each remote site.

Partitions and Calling Search Spaces

The dial plan for a multi-site IP WAN deployment with distributed call processing is similar to the dial plan for a single-site deployment. The only major difference is that the multi-site deployment can access remote sites across the IP WAN. Therefore, multi-site deployments usually require at least two route lists:

- One route list that always uses the PSTN for external (off-net) calls
- One route list that uses the IP WAN as the first-choice path for on-net calls and the PSTN as the second-choice path if the IP WAN is not available

Emergency Services

Emergency services are of great importance in the proper deployment of a voice system. This chapter presents a summary of the following major design considerations essential to planning for emergency calls:

This chapter presents some information specific to the 911 emergency networks as deployed in Canada and the United States. Many of the concepts discussed here are adaptable to other locales. Please consult with your local telephony network provider for appropriate implementation of emergency call functionality.

In the United States, some states have already enacted legislation covering the 911 functionality required for users in a multi-line telephone system (MLTS). The National Emergency Number Association (NENA) has also produced the *Technical Information Document on Model Legislation Enhanced 911 for Multi-Line Telephone Systems*, available online at

http://www.nena9-1-1.org/9-1-1TechStandards/TechInfoDocs/MLTS_ModLeg_Nov2000.PDF

The Federal Communications Commission (FCC) has also drafted a proposed new section to title 47, part 68, section 319, which is available at

http://www.apcointl.org/pbx/worddocs/mltspart68.doc

This chapter assumes that you are familiar with the generic 911 functionality available to residential PSTN users in North America. For more information on the subject, refer to the following URL for a description of the current state of E911 services in North America:

http://www.nena9-1-1.org/9-1-1%20Tutorial/9-1-1_tutorial.htm

Planning for 911 Functionality

This section highlights some of the functionality requirements for lifeline calls in multi-line telephone systems (MLTS). In the context of this section, lifeline calls are 911 calls services by the North American public switched telephone network (PSTN).

When planning an MLTS deployment, first establish all of the physical locations where phone services are needed. The locations can be classified as follows:

- Single building deployments, where all users are located in the same building.
- Single campus deployments, where the users are located in a group of buildings situated in close proximity.
- Multi-site deployments, where users are distributed over a wide geographical area and linked to the telephony call processing site via WAN connectivity.

The locations, or type of deployment, affect the criteria used to design and implement 911 services. The following sections describe the key criteria, along with typical and exceptional situations for each. When analyzing and applying these criteria, consider how they are affected by the phone service locations in your network.

Public Safety Answering Point (PSAP)

The public safety answering point (PSAP) is the party responsible for answering the 911 call and arranging the appropriate emergency response, such as sending police, fire, or ambulance teams. The physical location of the phone making the 911 call is the primary factor in determining the appropriate PSAP for answering that call. Generally, each building is serviced by one local PSAP.

To determine the responsible PSAP for a given location, contact a local public safety information service such as the local fire marshal or police department. Also, the phone directory of the local exchange carrier usually lists the agency responsible for servicing 911 calls in a given area.

Typical Situation

- For a given street address, there is only one designated PSAP.
- For a given street address, all 911 calls are routed to the same PSAP.

Exceptional Situation

- The physical size of the campus puts some of the buildings in different PSAP jurisdictions.
- Some of the 911 calls need to be routed to an on-net location (campus security, building security).

911 Network Service Provider

After identifying the responsible PSAPs, you must also identify the 911 network service providers to which each PSAP is connected. It is commonly assumed that PSAPs receive 911 phone calls from the PSTN, but that is not the case. Instead, 911 calls are carried over dedicated, regionally significant networks, and each PSAP is connected to one or more such regional networks. In the majority of cases, the incumbent Regional Bell Operating Company (RBOC) is the 911 network service provider for a PSAP. Some exceptions include military installations, university campuses, federal or state parks, or other locations where the public safety responsibility falls outside the jurisdiction of the local authorities and/or where a private network is operated by an entity other than a public local exchange carrier.

If you are in doubt about the 911 network service provider for a given PSAP, contacting the PSAP directly to verify the information.

Typical Situation

- For a given street address, the 911 network service provider is the local RBOC (more generally, the local exchange carrier). For a location served by Phone Company X, the corresponding PSAP is also served by Phone Company X.
- All 911 calls are routed directly to an off-net location, or all 911 calls are routed directly to an on-net location.

Exceptional Situation

- The local exchange carrier (LEC) through which the MLTS interfaces to the PSTN is *not* the same LEC that serves as 911 network service provider to the PSAP. (For example, the phone system is served by Phone Company X, but the PSAP is connected to Phone Company Y.) This situation might require either a special arrangement between the LECs or special, dedicated trunks between the phone system and the PSAP's 911 network service provider.
- Some LECs may not accept 911 calls on their networks. If this is the case, the only two options are to change LECs or to establish trunks (dedicated to 911 call routing) connected to an LEC that can route 911 calls to the appropriate PSAPs.
- Some (or all) of the 911 calls have to be routed to an on-net location such as campus security or building security. This situation can easily be accommodated during the design and implementation phases, but only if the destination of 911 calls for each phone has been properly planned and documented.

Interface Points into the Appropriate 911 Networks

For larger telephony systems, 911 connectivity might require many interface points. Typically, more than one E911 selective router is used within an LEC's territory, and these routers usually are *not* interconnected.

For example, an enterprise with a large campus could have the following situation:

- Building A located in San Francisco
- Building B located in San Jose
- San Francisco PD and San Jose PD are the appropriate PSAPs
- San Francisco PD and San Jose PD are served by the same 911 network service provider
- However, San Francisco PD and San Jose PD are served by different E911 selective routers operated by that same 911 network service provider!

This type of situation would require two separate interface points, one per E911 selective router. The information pertaining to the E911 selective router territories is generally kept by the incumbent LEC, and the local account representative for that LEC should be able to provide an enterprise customer with the pertinent information. Many LECs also provide the services of 911 subject matter experts who can consult with their own account representatives on the proper mapping of 911 access services.

Typical Situation

- For single-site deployments or campus deployments, there is usually only one PSAP for 911 calls.
- If access to only one PSAP is required, then only one interface point is required. Even if access to more than one PSAP is required, they might be reachable from the same E911 selective router, through the same centralized interface. If the enterprise's branch sites are linked via a WAN (centralized call processing), it is desirable to give each location its own local (that is, located inside each branch office) access to 911 to prevent 911 isolation during Survivable Remote Site Telephony (SRST) operation.

Exceptional Situation

- The physical size of the campus puts some of the buildings in different PSAP jurisdictions, *and*
- Some of the 911 calls have to be routed to different E911 selective routers, through different interface points.

Note Some of the information required to establish the geographical territories of PSAPs and E911 selective routers is available online or from various competitive local exchange carrier (CLEC) information web sites. (For example, https://clec.sbc.com/clec/hb/shell.cfm?section=782 provides some valuable data about the territory covered by SBC/Pacific Bell.) However, Cisco strongly recommends that you obtain proper written confirmation of the appropriate interface points from the LEC prior to the design and implementation phases of 911 call routing.

Interface Type

In addition to providing voice communications, the interfaces used to present 911 calls to the network must also provide identification data about the calling party.

Automatic Number Identification (ANI) refers to the E.164 number of the calling party, which is used by networks to route a 911 call to the proper destination. This number is also used by the PSAP to look up the Automatic Location Identification (ALI) associated with a call.

911 calls are source-routed, which means that they are routed according to the calling number. Even though different locations are all dialing the same number (911), they will reach different PSAPs based on their location, which is represented by the ANI (calling number).

You can implement 911 call functionality with either of the following interface types:

- Dynamic ANI assignment
- Static ANI assignment

While dynamic ANI assignment scales better (because it supports multiple ANIs) and lends itself to all but the smallest of applications, static ANI assignment can be used in a wider variety of environments, from the smallest to the largest systems.

Dynamic ANI (Trunk Connection)

The dynamic aspect of ANI refers to the fact that a system has many phones sharing access to the 911 network across the same interface, and the ANI transmitted to the network might need to be different for each call.

There are two types of dynamic ANI interfaces:

- Integrated Services Digital Network Primary Rate Interface (ISDN-PRI, or simply PRI)
- Centralized Automatic Message Accounting (CAMA).

PRI

This type of interface usually connects a telephony system to a PSTN Class 5 switch. The calling party number (CPN) is used at call setup time to identify the E.164 number of the calling party.

Most LECs treat the CPN differently when a call is made to 911. Depending upon the functionality available in the Class 5 switch and/or upon LEC or government policy, the CPN may not be used as the ANI for 911 call routing. Instead, the network may be programmed to use the listed directory number (LDN) or the bill-to number (BTN) for ANI purposes.

If the CPN is not used for ANI, then 911 calls coming from a PRI interface all look the same to the 911 network because they all have the same ANI, and they are all routed to the same destination (which might not be the appropriate one).

Some LECs offer a feature to provide CPN transparency through a PRI interface for 911 calls. With this feature, the CPN presented to the Class 5 switch at call setup is used for ANI to route the call. The feature name for this functionality varies, depending on the LEC. (For example, SBC calls it Inform 911 in California.)

Note The CPN *must* be a routable E.164 number, which means that the CPN must be entered in the database of the associated E911 selective router.

Note For Direct Inward Dial (DID) phones, the DID number could be used as the ANI for 911 purposes, but only if it is properly associated with an Emergency Service Number in the 911 service provider's network. For non-DID phones, use another number. (See Emergency Location Identification Number Mapping, page 8-7, for more information.)

Many Class 5 switches are connected to E911 selective routers through trunks that do not support more than one area code. In such cases, if PRI is used to carry 911 calls, then the only 911 calls that will be routed properly are those whose CPN (or ANI) have the same Numbering Plan Area (NPA) as the Class 5 switch.

Example

An MLTS is connected to a Class 5 switch in area code 514 (NPA = 514). If the MLTS were to send a 911 call on the PRI trunk, with a CPN of **450**.555.1212, the Class 5 switch would send the call to the E911 selective router with an ANI of **514**.555.1212 (instead of the correct **450**.555.1212), yielding inappropriate routing and ALI lookup.

To use PRI properly as a 911 interface, the system planner must ensure that the CPN will be used for ANI and must properly identify the range of numbers (in the format NPA NXX TNTN) acceptable on the link. For example, if a PRI link is defined to accept numbers within the range 514 XXX XXXX, then only calls that have a Calling Party Number with NPA = 514 will be routed appropriately.

CAMA

Centralized Automatic Message Accounting (CAMA) trunks also allow the MLTS to send calls to the 911 network, with the following differences from the PRI approach:

- CAMA trunks are connected directly into the E911 selective router. Extra mileage charges may apply to cover the distance between the E911 selective router and the MLTS gateway point.
- CAMA trunks support 911 calls only. The capital and operational expenses associated with the installation and operation of CAMA trunks support 911 traffic only.
- CAMA trunks for the MLTS market may be limited to a fixed area code, and the area code is typically implied (that is, not explicitly sent) in the link protocol. The connection assumes that all calls share the same deterministic area code, therefore only 7 or 8 digits are sent as ANI.

Note Cisco supports CAMA-based 911 functionality through the VIC-2CAMA trunk card.

Static ANI (Line Connection)

Static ANI provides a line (rather than a trunk) connection to the PSTN, and the ANI of the line is associated with all 911 calls made on that line, regardless to the CPN of the calling phone. A plain old telephone service (POTS) line is used for this purpose.

POTS lines are one of the simplest and most widely supported PSTN interfaces. A POTS line usually comes fully configured to accept 911 calls. In addition, the existing E911 infrastructure supports 911 calls from POTS lines very well.

The POTS approach has the following attributes:

- The operational costs associated with a POTS line are low.
- The POTS line can even serve as a backup line in case of power failure.
- The POTS line number can be used as the callback number entered into the ALI database.
- POTS lines represent the lowest cost 911 support for locations where user density does not justify local PRI or CAMA access into the PSTN.
- POTS lines are ubiquitous in PSTN installations.

All outgoing 911 calls through this type of interface are treated the same by the E911 network, and the tools that enable Cisco CallManager to control the ANI presented to the E911 network (such as calling party transformation masks) are irrelevant because the ANI can be only the POTS line number.

Emergency Response Location Mapping

The National Emergency Number Association (NENA) has recently proposed model legislation to be used by state and federal agencies in enacting the rules that govern 911 in enterprise telephony systems. One of the concepts in the NENA proposal is that of the emergency response location (ERL), which is defined as:

> *A location to which a 911 emergency response team may be dispatched. The location should be specific enough to provide a reasonable opportunity for the emergency response team to quickly locate a caller anywhere within it.*

Rather than having to identify each phone's location individually, the requirement allows for the grouping of phones into a "zone," the ERL. The maximum size of the ERL may vary, depending upon local implementation of the legislation, but we will use 7000 square feet (sq ft) as a basis for discussion in this section. (The concepts discussed here are independent of the maximum ERL size that may be allowed in any given state or region.)

An emergency location identification number (ELIN) is associated with each ERL. The ELIN is a fully qualified E.164 number, used to route the call within the E911 network. The ELIN is sent to the E911 network for any 911 call originating from the associated ERL. This process allows more than one phone to be the associated with the same fully qualified E.164 number for 911 purposes, and it can be applied to DID and non-DID phones alike.

Note This document does not attempting to present the actual requirements of any legislation. Rather, the information and examples presented here are for the purposes of discussion only. The system planner is responsible for verifying the applicable local requirements.

For example, assume a building has a surface area of 70,000 sq ft and 100 phones. In planning for 911 functionality, the building can be divided into 10 zones (ERLs) of 7000 sq ft each, and each phone can be associated with the ERL where it is located. When a 911 call is made, the ERL (which could be the same for multiple phones) is identified by sending the associated ELIN to the PSAP. If the phones were evenly distributed in this example, each group of 10 phones would have the same ERL and, therefore, the same ELIN.

The various legislations define a minimum number of phones (for example, 49) and a minimum surface area (for example, 40,000 sq ft) below which the requirements for MLTS 911 are not applicable. But even if the legislation does not require 911 functionality for a given enterprise, it is always best practice to provision for it.

Emergency Location Identification Number Mapping

In general, you must associate a single fully qualified E.164 number, know as the emergency location identification number (ELIN), with each ERL. (However, if using Cisco Emergency Responder, you can configure more than one ELIN per ERL.) The ELIN is used to route the call across the E911 infrastructure and is used by the PSAP as the index into the ALI database.

ELINs must meet the following requirements:

- They must be routable across the E911 infrastructure. (See the examples in the section on Interface Type, page 8-4.) If an ELIN is not routable, 911 calls from the associated ERL will, at best, be handled according to the default routing programmed in the E911 selective router.
- Once the ERL-to-ELIN mapping of an enterprise is defined, the corresponding ALI records must be established with the LEC so that the ANI and ALI database records serving the PSAP can be updated accurately.

The ELIN mapping process can be one of the following, depending on the type of interface to the E911 infrastructure for a given ERL:

- Dynamic ANI interface

  With this type of interface, the calling party number identification passed to the network is controlled by the MLTS. The telephony routing table of the MLTS is responsible for associating the correct ELIN with the call, based on the calling phone's ERL. Within Cisco CallManager, the calling party number can be modified by using transformation masks for calls made to 911. For example,

all phones located in a given ERL can share the same calling search space that lists a partition containing a translation pattern (911) and a calling party transformation mask that would replace the phone's CPN with the ELIN for that location.

- Static ANI interface

  With this type of interface, the calling party number identification passed to the network is controlled by the PSTN. This is the case if the interface is a POTS line. The ELIN is the phone number of the POTS line, and no further manipulation of the phone's calling party identification number is possible.

PSAP Callback

The PSAP might have to reach the caller after completion of the initial conversation. The PSAP's ability to call back relies on the information that it receives with the original incoming call.

The delivery of this information to the PSAP is a two-part process:

1. The Automatic Number Identification (ANI) is first sent to the PSAP. The ANI is the E.164 number used to route the call. In our context, the ANI received at the PSAP is the ELIN that the MLTS sent.
2. The PSAP then uses the ANI to query a database and retrieve the Automatic Location Identification (ALI). The ALI provides the PSAP attendant with information such as:
   - Caller's name
   - Address
   - Applicable public safety agency
   - Other optional information, which could include callback information. For example, the phone number of the enterprise's security service could be listed, to aid in the coordination of rescue efforts.

Typical Situation

- The ANI information is used for PSAP callback, which assumes that the ELINs are dialable numbers.
- The ELINs are PSTN numbers associated with the MLTS. If someone calls the ELIN from the PSTN, the call will terminate on an interface controlled by the MLTS.
- It is the responsibility of the MLTS system administrator to program the call routing so that calls made to any ELIN in the system will ring a phone (or multiple phones) in the immediate vicinity of the associated ERL.
- Once the ERL-to-ELIN mapping is established, it needs be modified only when there are changes to the physical situation of the enterprise. If phones are simply added, moved, or deleted from the system, the ERL-to-ELIN mapping and its associated ANI/ALI database records need not be changed.

Exceptional Situation

- Callback to the immediate vicinity of the originating ERL may be combined with (or even superseded by) routing the callback to an on-site emergency desk, which will assist the PSAP in reaching the original caller and/or provide additional assistance with the emergency situation at hand.
- The situation of the enterprise could change, for example, due to area code splits, city or county service changes requiring a new distribution of the public safety responsibilities, new buildings being added, or any other change that would affect the desired routing of a call for 911 purposes. Any of these evens could require changes in the ERL-to-ELIN mapping and the ANI/ALI database records for the enterprise.

Nomadic Phone Considerations

All discussions in this chapter thus far have relied upon the assumption that the phone locations are static. If, however, phones are moved across ERL boundaries, then 911 calls from the newly relocated phone will not be routed correctly. Because it is now physically located in a different ERL, the phone should use the ELIN of its current ERL. If the configuration is not changed in the Cisco CallManager database, the following events will occur:

- The ELIN of the previous ERL will be used to route calls on the E911 infrastructure.
- The egress point from the IP network to the E911 infrastructure might be incorrect.
- The callback functionality provided to the PSAP might reach the wrong destination.
- The ALI information provided to the PSAP might result in the dispatching of emergency response personnel to the wrong location.
- The location-based call admission control for the phone might not properly account for the WAN bandwidth usage of the phone, yielding possible over-subscription or under-subscription of WAN bandwidth resources.

The only way to remedy this situation is to manually update the phone's configuration (including its calling search space and location information) in Cisco CallManager to reflect its new physical location.

Cisco Emergency Responder

Ease of administration for moves, adds, and changes (MACs) is one of the key advantages of Voice over IP (VoIP) technology. To provide for MACs that automatically update 911 information without user intervention, Cisco has developed a product called the Cisco Emergency Responder (Cisco ER).

Cisco ER provides the following primary functionality:

- Dynamic association of a phone to an ERL, based on the detected physical location of the phone.
- Dynamic association of the ELIN to the calling phone, for callback purposes. In contrast to non-ER scenarios outlined in preceding sections, Cisco ER enables the callback to ring the exact phone that initiated the 911 call.
- On-site notification to designated parties (by pager, email, or phone call) to inform them that there is an emergency call in progress. The notification includes the calling party name and number, the ERL, and the date and time details associated with the call.

For more information on Cisco ER, refer to the section on Cisco Emergency Responder Considerations, page 8-13, and to the Cisco ER product documentation available online at

http://www.cisco.com/univercd/cc/td/doc/product/voice/respond/index.htm

The key functionality of Cisco ER relies on the detection of the phone's location by discovery of the network port (Layer 2 port, such as a Fast Ethernet switched port) from which the phone made the 911 call. The discovery mechanism relies on two main assumptions:

- The wired infrastructure of the enterprise is well established and does not change sporadically.
- The infrastructure is available for Cisco ER to browse; that is, Cisco ER can establish Telnet or Simple Network Management Protocol (SNMP) sessions to the underlying network infrastructure and can scan the network ports for the discovery of connected phones.

Once Cisco ER discovers the originating port for the call, it associates the call with the pre-established ERL for the location of that port. This process also yields an association with a pre-established ELIN for the location and the selection of the appropriate egress point to the E911 infrastructure, based on the originating ERL.

With Cisco ER, the ERL-to-ELIN mapping process described in the preceding sections still applies, but with one variation: without Cisco ER, each ERL is associated with only one ELIN, but Cisco ER allows for the use of two or more ELINs per ERL. The purpose of this enhancement is to cover the specific case of more than one 911 call originating from a given ERL within the same general time period, as illustrated by the following examples.

Example 1

- Phone A and phone B are both located within ERL X, and ERL X is associated with ELIN X.
- Phone A makes a 911 call at 13:00 hours. ELIN X is used to route the call to PSAP X, and PSAP X answers and releases the call. Then, at 13:15 hours, phone B makes a 911 call. ELIN X is again used to route the call to PSAP X.
- PSAP X, after releasing the call from phone B, decides to call back phone A for further details pertaining to phone A's original call. The PSAP dials ELIN X, and gets phone B (instead of the desired phone A).

To work around this situation, Cisco ER allows you to define a pool of ELINs for each ERL. This pool provides for the use, in a round-robin fashion, of a distinct ELIN for each successive call. With the definition of two ELINs for ERL X in our example, we now have the situation described in Example 2.

Example 2

- Phone A and phone B are both located within ERL X. ERL X is associated with both ELIN X1 and ELIN X2.
- Phone A makes a 911 call at 13:00 hours. ELIN X1 is used to route the call to PSAP X, and PSAP X answers and releases the call. Then, at 13:15 hours, phone B makes a 911 call, and ELIN X2 is used to route this call to PSAP X.
- PSAP X, after releasing the call from phone B, decides to call back phone A for further details pertaining to phone A's original call. The PSAP dials ELIN X1 and gets phone A.

Of course, if a third 911 call were made and there were only two ELINs for the ERL, the situation would again be similar to Example 1.

Emergency Call String

It is highly desirable to configure a dial plan so that the system easily recognizes emergency calls, whether an access code (for example, 9) is used or not. The emergency string for North America is generally 911. Cisco strongly recommends that you configure the system to recognize both the strings 911 and 9911.

Cisco also strongly recommends that you explicitly mark the emergency route patterns with Urgent Priority so that Cisco CallManager does *not* wait for the inter-digit timeout (Timer T.302) before routing the call.

Other emergency call strings may be supported concurrently on your system. Cisco highly recommends that you provide your telephony system users with training on the selected emergency call strings.

Also, it is highly desirable that users be trained to react appropriately if they dial the emergency string by mistake. In North America, 911 may be dialed in error by users trying to access a long distance number through the use of 9 as an access code. In such a case, the user should remain on the line to confirm that there is no emergency, and therefore no need to dispatch emergency personnel.

Gateway Considerations

Consider the following factors when selecting the gateways to handle emergency calls for your system:

Gateway Placement

Within the local exchange carrier (LEC) networks, 911 calls are routed over a locally significant infrastructure based on the origin of the call. The serving Class 5 switches are connected either directly to the relevant PSAP for their location or to an E911 selective router, which itself is connected to a group of PSAPs significant for its region.

With an IP-based enterprise telephony architecture such as Cisco AVVID, it is possible to route calls on-net to gateways that are remotely situated. As an example, a phone located in San Francisco could have its calls carried over an IP network to a gateway situated in San Jose, and then sent to the LEC's network.

For 911 calls, it is critical to chose the egress point to the LEC network so that emergency calls are routed to the appropriate local PSAP. In the example above, a 911 call from the San Francisco phone, if routed to a San Jose gateway, could not reach the San Francisco PSAP because the San Jose LEC switch receiving the call does not have a link to the E911 selective router serving the San Francisco PSAP. Furthermore, the San Jose area 911 infrastructure would not be able to route the call based on a San Francisco calling party number.

As a rule of thumb, route 911 calls to a gateway physically co-located with the originating phone. Contact the LEC to explore the possibility of using a common gateway to aggregate the 911 calls from multiple locations.

Gateway Blocking

It is highly desirable to protect 911 calls from "all trunks busy" situations. If a 911 call needs to be connected, it should be allowed to proceed even if other types of calls are blocked due to lack of trunking resources. To provide for such situations, you can dedicate an explicit trunk group just for 911 calls.

It is acceptable to route emergency calls exclusively to an emergency trunk group. Another approach is to send emergency calls to the same trunk group as the regular PSTN calls (if the interface permits it), with an alternative path to a dedicated emergency trunk group. This latter approach allows for the most flexibility.

As an example, we can point emergency calls to a PRI trunk group, with an alternate path (reserved exclusively for emergency calls) to POTS lines for overflow conditions. If we put 2 POTS lines in the alternate trunk group, we are guarantying that a minimum of two simultaneous 911 calls can be routed in addition to any calls that were allowed in the main trunk group.

If the preferred gateway becomes unavailable, it may be acceptable to overflow emergency calls to an alternate number so that an alternate gateway is used. For example, in North America calls dialed as 911 could overflow to an E.164 (non-911) local emergency number. This approach does not take advantage

of the North American 911 network infrastructure (that is, there is no selective routing, ANI, or ALI services), and it should be used only as an last resort to avoid blocking the emergency call due to a lack of network resources.

Answer Supervision

Under normal conditions, calls made to an emergency number should return answer supervision upon connection to the PSAP. The answer supervision may, as with any other call, trigger the full duplex audio connection between the on-net caller and the egress interface to the LEC's network.

With some North American LECs, answer supervision might not be returned when a "free" call is placed. This may be the case for some toll-free numbers (for example, 800 numbers). In exceptional situations, because emergency calls are considered "free" calls, answer supervision might not be returned upon connection to the PSAP. You can detect this situation simply by making a 911 test call. Upon connection to the PSAP, if audio is present, the call timer should record the duration of the ongoing call; if the call timer is absent, it is very likely that answer supervision was not returned. If answer supervision is not returned, Cisco highly recommends that you contact the LEC and report this situation because it is most likely not the desired functionality.

If this situation cannot be rectified by the Local Exchange Carrier, it would be advisable to configure the egress gateway *not* to require answer supervision when calls are placed to the LEC's network, and to cut through the audio in both directions so that progress indicator tones, intercept messages, and communications with the PSAP are possible even if answer supervision is not returned.

By default, Cisco IOS-based H.323 gateways must receive answer supervision in order to connect audio in both directions. To forego the need for answer supervision on these gateways, use the following commands:

- **progress_ind alert enable 8**

  This command provides the equivalent of receiving a progress indicator value of 8 (in-band information now available) when alerting is received. This command allows the POTS side of the gateway to connect audio toward the origin of the call.

- **voice rtp send-recv**

  This command allows audio cut-through in both the backward and forward directions before a Connect message is received from the destination switch. This command affects all Voice over IP (VoIP) calls when it is enabled.

In all cases, Cisco highly recommends that you test 911 call functionality from all call paths and verify that answer supervision is returned upon connection to the PSAP.

Cisco Emergency Responder Considerations

Device mobility brings about special design considerations for emergency calls. Cisco Emergency Responder (ER) can be used to track device mobility and to adapt the system's routing of emergency calls based on a device's dynamic physical location.

Device Mobility Across Call Admission Control Locations

In a centralized call processing deployment, Cisco ER cannot fully support device movement across call admission control locations because Cisco CallManager does not know about device movements. For example, if you physically move a phone from Branch A to Branch B but the phone's call admission control location remains the same (for example, Location_A), then it is possible that calls made to 911 from that phone would be blocked due to call admission control denial if all available bandwidth to Location_A is in use for other calls. This call blocking occurs even if the phone, now in location B, is physically co-located with the gateway used to connect to the PSAP for location B.

For the same reasons, Cisco ER cannot support device movement across gatekeeper-controlled call admission control zones. However, Cisco ER can fully support device movements within those locations, where no call admission control is required.

In centralized call processing deployments, Cisco ER automatically supports device movement within branches. However, if a device is moved between branches, manual intervention is required to adapt the device's location and region parameters before Cisco ER can fully support 911 calls.

Default Emergency Response Location

If Cisco ER cannot directly determine the physical location of a phone, it assigns a default emergency response location (ERL) to the call. The default ERL points all such calls to a specific PSAP. Although there is no universal recommendation as to where calls should be sent when this situation occur, it is usually desirable to choose a PSAP that is centrally located and that offers the largest public safety jurisdiction. It is also advisable to populate the ALI records of the default ERL's emergency location identification numbers (ELINs) with contact information for the enterprise's emergency numbers and to offer information about the uncertainty of the caller's location. In addition, it is advisable to mark those ALI records with a note that a default routing of the emergency call has occurred.

Soft Clients

In cases where soft clients such as Cisco IP SoftPhone are used within the enterprise, Cisco ER can provide device mobility support. However, if the soft client is used outside the boundaries of the enterprise (for example, VPN access from a home office or hotel), Cisco ER will not be able to determine the location of the caller. Furthermore, it is unlikely that the Cisco AVVID system would have a gateway properly situated to allow sending the call to the appropriate PSAP for the caller's location.

It is a matter of enterprise policy to allow or disallow the use of soft clients for 911 calls. It is highly advisable to disallow 911 calls by policy for soft clients that can roam across the internet. Nevertheless, if such a user were to call 911, the best-effort system response would be to route the call to either an on-site security force or a large PSAP close to the system's main site.

The following paragraph is an example notice that you could issue to users to warn them that emergency call functionality is not guaranteed to soft client users:

> *Emergency calls should be placed from phones that are located at the site for which they are configured (for example, your office). A local safety authority might not answer an emergency call placed from a phone that has been removed from its configured site. If you must use this phone for emergency calls while away from your configured site, be prepared to provide the answering public safety authority with specific information regarding your location. Use a phone that is locally configured to the site (for example, your hotel phone or your home phone) for emergency calls when traveling or telecommuting.*

Test Calls

For any enterprise telephony system, it is a good idea to test 911 call functionality, not only after the initial installation, but regularly, as a preventive measure.

The following suggestions can help you carry out the testing:

- Contact the PSAP to ask for permission before doing any tests, and provide them with the contact information of the individuals making the tests.
- During each call, indicate that it is *not* an actual emergency, just a test.
- Confirm the ANI and ALI that the call taker has on their screen.
- Confirm the PSAP to which the call was routed.
- Confirm that answer supervision was received by looking at the call duration timer on the IP phone.

PSAP Callback to Shared Directory Numbers

Cisco ER handles the routing of inbound calls made to emergency location identification numbers (ELINs). In cases where the line from which a 911 call was made is a shared directory number, the PSAP callback will cause all shared directory number appearances to ring. Any of the shared appearances can then answer the call, which means that it may not be the phone from which the 911 call originated.

Voice Mail Integration

This chapter summarizes design considerations for integrating the following types of voice mail systems with Cisco CallManager:

Integrating Third-Party Voice Mail Systems

Cisco CallManager Release 3.2 and later can support a maximum of four third-party voice mail systems per cluster. You can integrate the following types of voice mail systems with Cisco CallManager through Voicemail Profiles, which you can configure on a per-line basis:

SMDI-Capable Voice Mail Systems

Cisco CallManager fully supports the BellCore/Telcordia Simplified Message Desk Interface (SMDI) protocol through either the Cisco Messaging Interface (CMI) service, which runs on a server (ideally, the publisher) within a cluster, or the Cisco VG248 Analog Phone Gateway. If you use the VG248, the VG248 itself (and not the Cisco CallManager server) provides the SMDI link directly.

Non-SMDI Serial-Capable Voice Mail Systems

The Cisco VG248 with software Release 1.2 and later also supports NEC's Message Center Interface (MCI) as well as Ericsson's proprietary serial protocol. This capability enables you to place the VG248 between the PBX and the voice mail system, thereby enabling a dual integration with Cisco CallManager.

The voice path can be delivered in a number of different ways, but for serial-based integrations it is typically derived from a suitable Skinny Client Control Protocol (SCCP) or Media Gateway Control Protocol (MGCP) gateway (such as the Cisco WS-X6608 or VG248) that can provide analog FXS ports. The VG248 is generally recommended because of its better cost model and because it does not require a slot within a Catalyst 6500 Series chassis.

Voice Mail Integration Using Cisco DPA

The Cisco Digital PBX Adapter (DPA) emulates both digital PBX ports as well as digital telephones, and it is available in two versions:

- DPA 7630 for Lucent/Avaya G3 — emulates the Lucent 7405 phone
- DPA 7610 for Nortel Meridian 1 — emulates the Nortel 2616 phone

Both products are designed specifically for the Avaya/Octel 200/300 Serenade and the 250/350 Aria voice mail systems and can be deployed in either single or dual integration mode.

A single voice mail system such as the Avaya/Octel 350 can be shared between multiple Cisco CallManager clusters when using the DPA. In that case, the message waiting indicator (MWI) ports are essentially daisy-chained between DPAs.

Integrating Cisco Unity

Cisco Unity can be deployed in either Unified Messaging or Voicemail-only mode. This functionality has no bearing on the physical method of integration to Cisco CallManager and/or the PBX.

Cisco Unity integrates to Cisco CallManager by emulating IP phones, namely the Cisco IP Phone 7960. In this way, the solution is completely IP.

A single Unity system can also support multiple Cisco CallManager clusters.

Unity can also operate in a dual integration mode (that is, simultaneous integration to both Cisco CallManager and a PBX). This mode can help mitigate migration issues.

For a complete list of supported PBXs, refer to the Unity product documentation, available at

http://cisco.com/go/interoperability

CHAPTER 10

Directory Access and Integration

This chapter summarizes design considerations for providing Cisco IP Telephony endpoints, such as Cisco IP Phones and Cisco IP SoftPhone, with access to a corporate Lightweight Directory Access Protocol (LDAP) directory. It also covers the main design principles for integrating Cisco CallManager with a corporate LDAP directory. The main topics include:

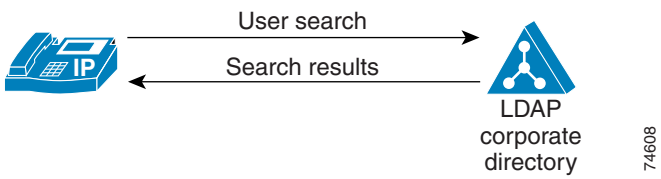
Directory Access Versus Directory Integration

The following definitions and distinctions apply throughout this document:

- *Directory access* refers to the ability of Cisco IP Telephony endpoints, such as Cisco IP Phones and Cisco IP SoftPhone, to access a corporate Lightweight Directory Access Protocol (LDAP) directory.
- *Directory integration* refers to the ability of an application, such as Cisco CallManager, to store its user-related information in a centralized corporate LDAP directory instead of using its own embedded database.

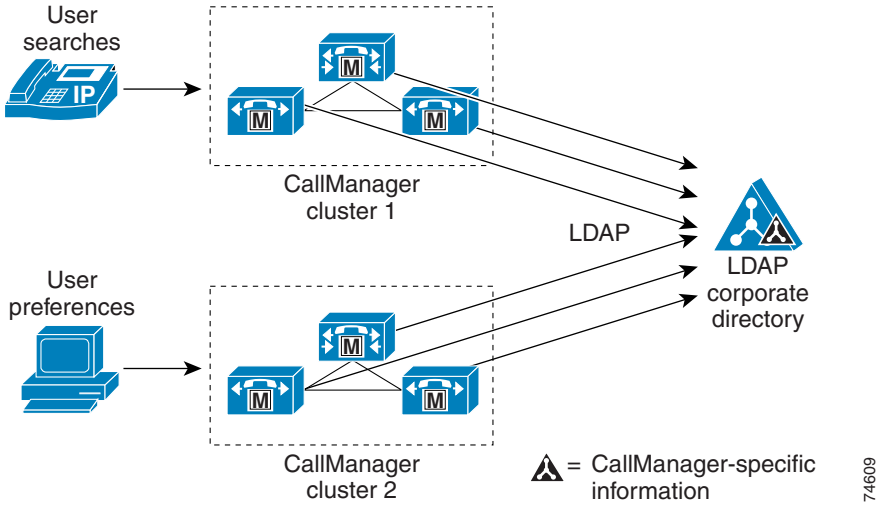
Figure 10-1 illustrates *directory access* as it is defined in this chapter. In this example, the access is provided to a Cisco IP Phone. The client application performs a user search against an LDAP directory, such as the corporate directory of an enterprise, and receives a number of matching entries. In this example, one entry can be selected and used to dial the corresponding person from the Cisco IP Phone.

Figure 10-1 Example of Directory Access for a Cisco IP Phone



By contrast, *directory integration* of several applications with a corporate directory means that these applications actually store their user-related information in a centralized directory instead of using their own embedded databases. Figure 10-2 depicts an example of directory integration as it is defined in this chapter.

Figure 10-2 Example of Directory Integration with Cisco CallManager



By default, Cisco CallManager stores user information (such as which devices the user controls, speed dial numbers configured on the IP phone, and so on) in an embedded LDAP directory. However, Cisco CallManager can also be integrated with a corporate LDAP directory, which is normally used to store general employee information such as email address, office address, and job title. In those cases, Cisco CallManager no longer uses its own embedded directory but stores its specific user information in the corporate directory.

Note As of Cisco CallManager Release 3.3(2), directory integration is supported for Microsoft Active Directory and Sun/iPlanet/Netscape Directory Server releases 4.*x* and 5.1.

Integrating applications such as Cisco CallManager with a corporate directory also has the following implications, which go beyond simply providing directory access to endpoints:

- The directory schema must be extended to store the application-specific user attributes in the corporate directory. This operation is not trivial and requires a good knowledge of the directory structure.
- The applications must be able to contact the directory at all times. Availability of the directory service can impact application functionality.
- Additional load is introduced on the directory, in terms of both data storage and read/write queries. Careful planning and sizing is recommended to avoid oversubscription of the servers.

While there are numerous advantages to directory integration across applications, it is important to understand all of its implications and to verify the business needs for each specific deployment.

Directory Access for Cisco IP Telephony Endpoints

The guidelines contained in this section apply regardless of whether Cisco CallManager and other IP telephony applications have been integrated to a corporate directory or not. The end-user perception in both cases is the same because the differences affect only the way applications store their user information and how such information is kept consistent across the network.

The following sections summarize how to configure corporate directory access to any LDAP v3-compliant directory server for the following Cisco IP Telephony endpoints:

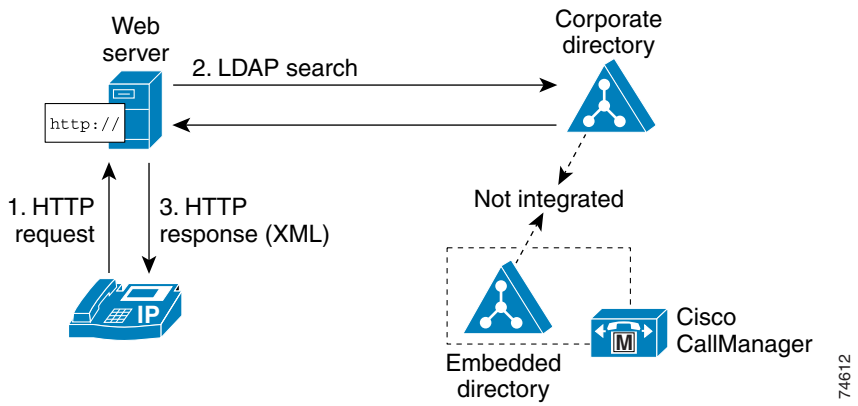
- Cisco IP Phones (7940 and 7960)
- Cisco IP SoftPhone

Directory Access for Cisco IP Phones

The Cisco IP Phones 7940 and 7960 can search a corporate LDAP directory when a user activates the Directories button on the phone. The IP Phones use the Hyper Text Transfer Protocol (HTTP) to send requests to a web server. The responses from this server must contain some specific eXtensible Markup Language (XML) objects that can be interpreted and displayed by the phone. In the case of a corporate directory search, the web server operates as a proxy by receiving the request from the phone and translating it into an LDAP request, which is in turn sent to the corporate directory server. The response is then interpreted and sent back to the phone, after having been encapsulated in the appropriate XML objects.

Figure 10-3 illustrates this mechanism in a deployment where Cisco CallManager has not been integrated with the corporate directory. Note that, in this scenario, Cisco CallManager is not involved in the message exchange.

Figure 10-3 Directory Access Mechanism for Cisco IP Phones



The proxy function provided by the web server can be configured using the Cisco IP Phone Services Software Development Kit (SDK) version 2.0, which includes the Cisco LDAP Search COM Server. When you install the SDK, the Cisco LDAP Search COM Server is installed automatically. (Refer to the SDK product documentation for more details on the installation procedure.)

After installing the SDK, you must customize the Active Server Pages (ASPs) served by the server. Some sample scripts are provided as part of the SDK. By editing these scripts, you can provide a wide range of directory searches according to the specific needs of your deployment.

Once you have customized the ASPs, set the URL Directories field within Cisco CallManager's Enterprise Parameters to point to the web server mentioned above, then reset the Cisco IP Phones.

Directory Access for Cisco IP SoftPhone

As of Release 1.2, Cisco IP SoftPhone has a built-in mechanism to access and search LDAP directories. Refer to the product documentation for details on how to configure this feature.

Directory Integration with Cisco CallManager

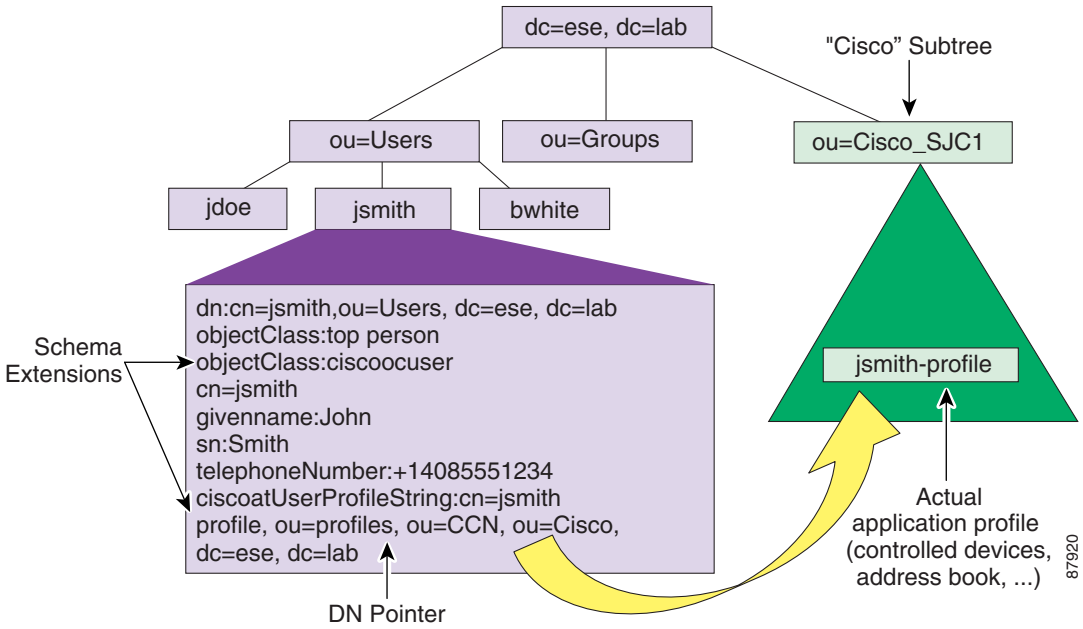
Cisco CallManager uses an embedded Microsoft SQL database to store system and device configuration data, such as dial plan information, IP phone and gateway configuration, and media resource utilization. It also uses an embedded LDAP directory to store user and application profiles, such as devices controlled by a user, speed dial configurations, and personal address books.

Both the SQL database and the LDAP directory run on each Cisco CallManager server within a cluster, and replication agreements are automatically set up between servers. The publisher server contains the master copy of both the SQL database and the LDAP directory, and it handles replication to all subscriber servers, which contain read-only copies of both repositories.

In order to store application-specific information in an LDAP directory, Cisco CallManager adopts an approach that is valid both when using the embedded directory and when integrating with a corporate directory.

Since different directory vendors typically use different User object models with several additional, non-standard attributes, Cisco CallManager assumes that the User object contains only the standard LDAPv3 core attributes. The User object is then augmented with an auxiliary class, ciscoocUser, which contains a single attribute called ciscoatUserProfileString. (Earlier versions of Cisco CallManager used a different attribute called ciscoatUserProfile. For backward compatibility and compatibility with other Cisco applications, both attributes are maintained in the schema.) This attribute is a Distinguished Name pointer to another object in the directory, which contains the application-specific profile. With this approach, the impact on the core User object is minimized, and all the application-specific information can be stored in a separate Organizational Unit (OU) within the directory, usually referred to as the Cisco subtree or the Cisco DIT. Figure 10-4 shows a graphical representation of this process.

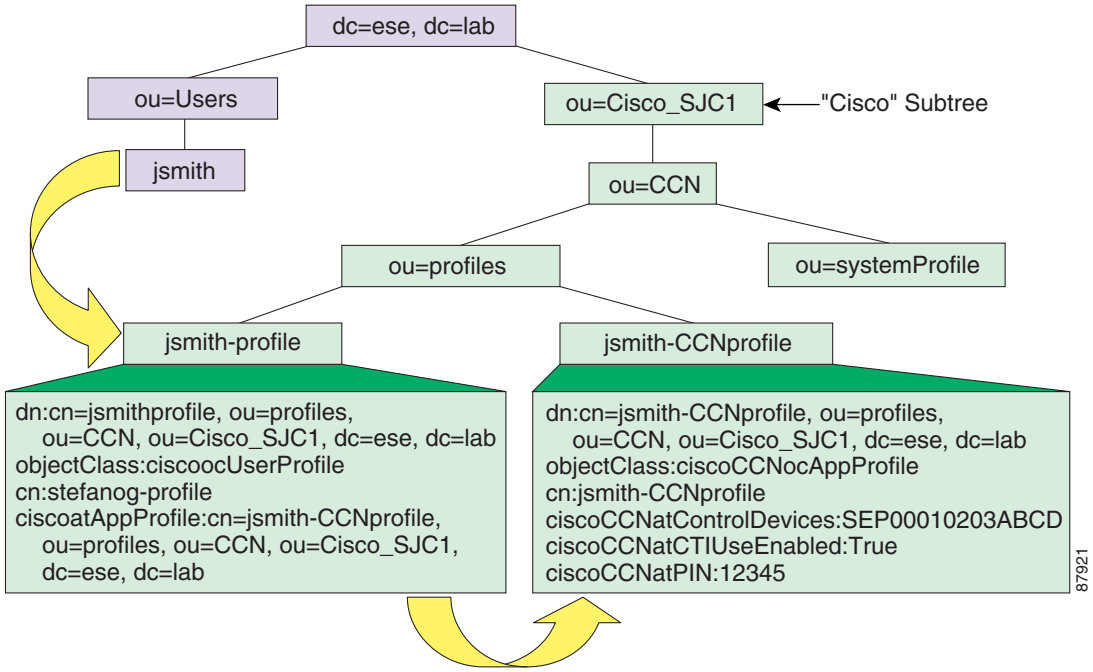
Figure 10-4 Cisco CallManager Directory Approach



The object pointed to by the ciscoatUserProfileString attribute belongs to another auxiliary class called ciscoocUserProfile. Its canonical name (CN) is obtained by adding the suffix “-profile” to the CN of the core User object. The main purpose of this object is to store other pointers to the specific profile objects for all Cisco applications that integrate with the directory. (This is done through the multi-valued

attribute named ciscoatAppProfile.) The application profile used by Cisco CallManager belongs to the auxiliary class called ciscoCCNocAppProfile, and its CN is obtained by adding the suffix "-CCNprofile" to the CN of the core user. This is where Cisco CallManager stores the user's Extension Mobility PIN number, the list of devices controlled by this user, whether this user is permitted to use CTI applications, and so on. Figure 10-5 depicts the relationship between these two profile objects.

Figure 10-5 Relationship Between Core User Object, User Profile, and Application Profile



To integrate Cisco CallManager with an external LDAP directory, run the Cisco Customer Directory Integration Plugin, which is bundled with the Cisco CallManager software. This plugin serves three main purposes:

- It extends the corporate directory schema to accommodate the application-specific objects and attributes.
- It populates the "Cisco" subtree with the configuration objects needed by Cisco CallManager.
- It configures Cisco CallManager to use the corporate directory and disables its embedded directory.

For more details on how to install the Cisco Customer Directory Integration Plugin, refer to the product documentation.

Best Practices

The following best practices apply when integrating Cisco CallManager with a corporate directory such as Microsoft Active Directory (AD):

- Ensure that the integration is planned and implemented by your organization's directory team.
- Before integration, test in a lab setup against an exact replica of the production AD.
- Back up the AD forest prior to integration.
- In the Cisco Customer Directory Integration Plugin configuration, use domain names resolvable by Domain Name System (DNS) instead of specific AD server names.

- Use Cisco IOS Server Load Balancing (SLB) on a Catalyst 6000 if AD-integrated DNS load balancing is not available.
- Set the "User Search base" to the lowest point in the tree that contains all required users.
- If deploying Cisco CallManager Release 3.3, make sure the "User Creation base" is contained within the configured "User Search base."
- Make sure that a Microsoft Windows 2000 Domain Controller (DC) is local to each Cisco CallManager server deployed, and that the directory infrastructure is highly available.
- Manage AD user accounts by using Active Directory Users & Computers (ADU&C), which is part of the Microsoft Management Console (MMC).
- Manage Cisco CallManager attributes by using Cisco CallManager Administration.
- Set passwords from ADU&C or Windows PC.
- Use a dedicated "service account" for Cisco CallManager to access the directory, and grant the minimum rights necessary for this account.
- With new deployments, try to plan for AD integration from the beginning.

The following limitations also apply when integrating Cisco CallManager with a corporate directory such as Microsoft Active Directory (AD):

- Currently, no data migration is performed by the Cisco Customer Directory Integration Plugin. Sample migration scripts are available as guidelines through your Cisco account team.
- Prior to Cisco CallManager Release 3.3, multiple cluster deployments could not have overlapping search bases within an AD tree. Integration of multiple Cisco CallManager clusters with the same corporate directory should be reviewed by a Cisco account team.
- A Cisco CallManager cluster can be integrated with a single tree within an AD forest.
- AD passwords cannot be set or reset from Cisco CallManager User or Cisco CallManager Administration pages.

IP Phone Services

Cisco IP Phone Services are applications that utilize the web client capabilities on the Cisco IP Phones 7940 and 7960. The Cisco IP Phone firmware contains a micro-browser that enables limited web browsing capability. For purposes of this chapter, the term *phone service* refers to an application that transmits and receives content to and from the Cisco IP Phone.

An IP Phone service can be initiated in several ways:

- User initiated — An IP Phone user presses the Services button, which sends an HTTP GET message to Cisco CallManager for displaying a list of user-subscribed phone services.
- Phone initiated — An IDLE URL timeout can be set within the IP Phone firmware. When this timeout value is exceeded, the IP Phone firmware itself initiates an HTTP GET to the IDLE URL location specified in the phone's enterprise parameter setting.
- Phone Service initiated — The phone service application can "push" content to the IP Phone via an HTTP POST message to the IP Phone.

Figure 11-1 illustrates these methods.

Figure 11-1 Methods of Initiating IP Phone Services

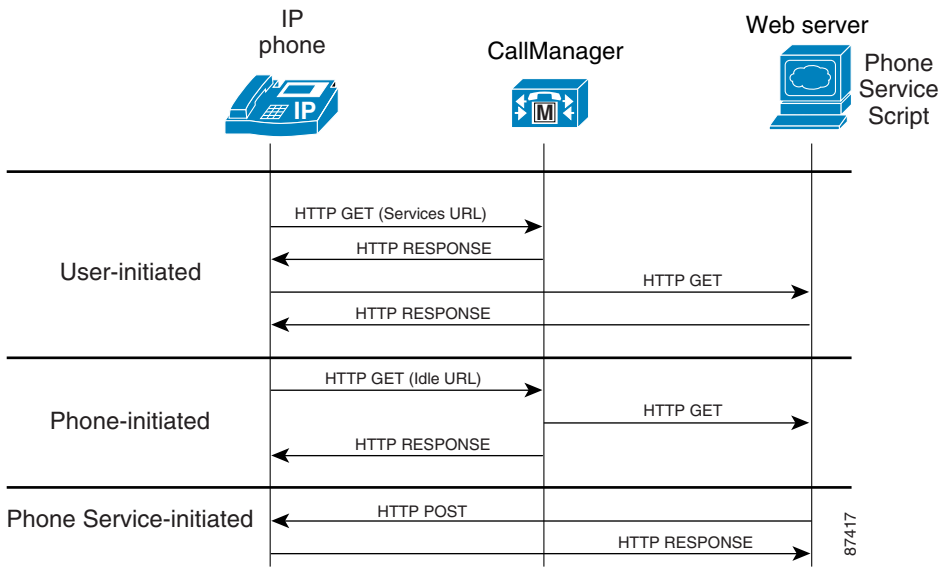


Table 11-1 lists the various URLs that you can set for the IP Phones.

Table 11-1 IP Phone URL Settings

Setting	Description	Comments
URL Information	Displays on-phone help of the phone keys when the user presses the "i" button located to the right of the dial keypad	Points to the GetTelecasterHelpText.asp file located on the Cisco CallManager configured for the IP Phone
URL Services	Retrieves the list of user-subscribed phone services	Points to the getservicesmenu.asp file located on the Cisco CallManager configured for the IP Phone
URL Directories	Retrieves a user directory search list	Points to the xmldirectory.asp file located on the Cisco CallManager configured for the IP Phone
URL Messages	Redirects the IP Phone to the set URL location when a user presses the Messages button	By default, the Messages button is configured for connection to voice mail. This field is usually left blank.
URL Idle	Provides content delivery to the phone display when the phone is in IDLE state	Points to an administrator-configured URL when the IP Phone has not been used for the time specified in the Idle Time field
URL Proxy	Specifies the host and port used to proxy HTTP requests for access to a non-local host address from the phone's HTTP client	

The IP Phones can process a limited set of Cisco-defined eXtensible Markup Language (XML) objects for enabling the user interface (UI) between the IP Phone and the web server that contains the running phone service. Figure 11-2 illustrates the message flow across this interface.

Figure 11-2 Basic Message Flow for Cisco IP Phone Services

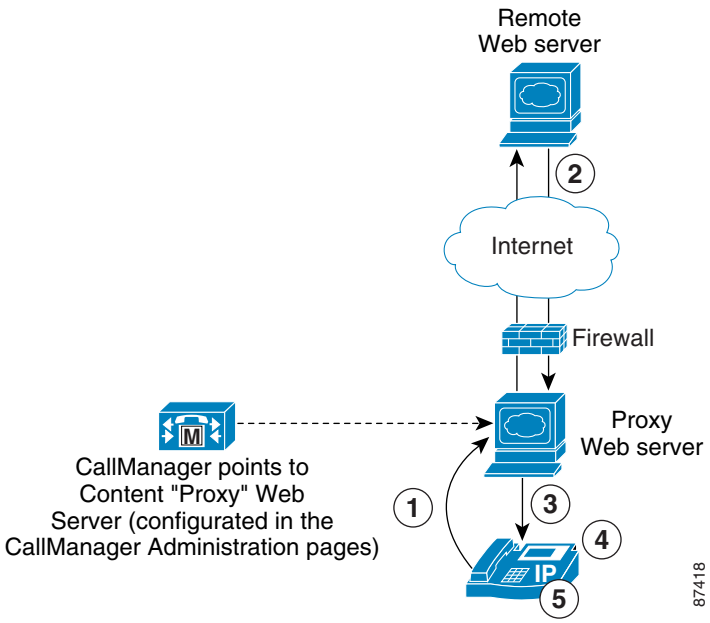


Figure 11-2 shows the following message flow sequence:

1. Cisco IP Phone HTTP Client performs an HTTP GET for a specified URL.
2. HTTP web server processes the request and formats the data returned.
3. HTTP web server returns the HTTP response of XML objects or plain text to the IP Phone.
4. IP Phone parses the HTTP response header for ContentType of “text” or XML.
5. IP Phone presents data and options to the user according to the server response.

The Cisco CallManager server does not execute the actual phone service. This service resides on a separate web server, which is usually behind a corporate firewall for security purposes. That server can also proxy to a remote server to obtain Internet content.

Note IP Phone services must reside on a separate web server. Running phone services on the Cisco CallManager server is not supported.

Integration Considerations

The Cisco IP Phone Services application is, for the most part, an HTTP client. It uses Cisco CallManager only as a redirect server to the location of the subscribed service. This section presents some design considerations for integrating IP Phone services with Cisco CallManager.

Scalability

Because Cisco CallManager acts as a redirect server to the phone service, there is minimal performance impact on Cisco CallManager when a user initiates a phone service request by pressing the Services key.

Because the IP Phone is either an HTTP client or server, estimating required bandwidth used by an IP Phone service is similar to estimating the bandwidth of an HTTP browser accessing the same text as HTTP content residing on the web hosting server.

Security

Existing IP telephony security design guidelines recommend that the IP Phones remain on a separate VLAN from the data stores. With this guideline in mind, there are several options for configuring IP Phone services:

- The proxy web server running the phone services can reside on the voice VLAN. This configuration is easier to implement, and access filters can be set from the proxy web server only.

  The phone services should reside on their own dedicated server on the voice VLAN. Performance is slower for streaming media because the proxy web server has to store all of the media. If the content comes from the Internet, then the media might have to be buffered first.

- The proxy web server running the phone services can reside on the data VLAN.

Any IP Phone service traffic must be tagged to allow it to access any secured network segment outside the voice VLAN. One way to enable this access is by applying port access filters at a router or firewall that separates the voice and data VLAN segments. Table 11-2 lists some of the relevant ports that an IP Phone service can use.

Table 11-2 Common Ports Used by Phone Service Applications

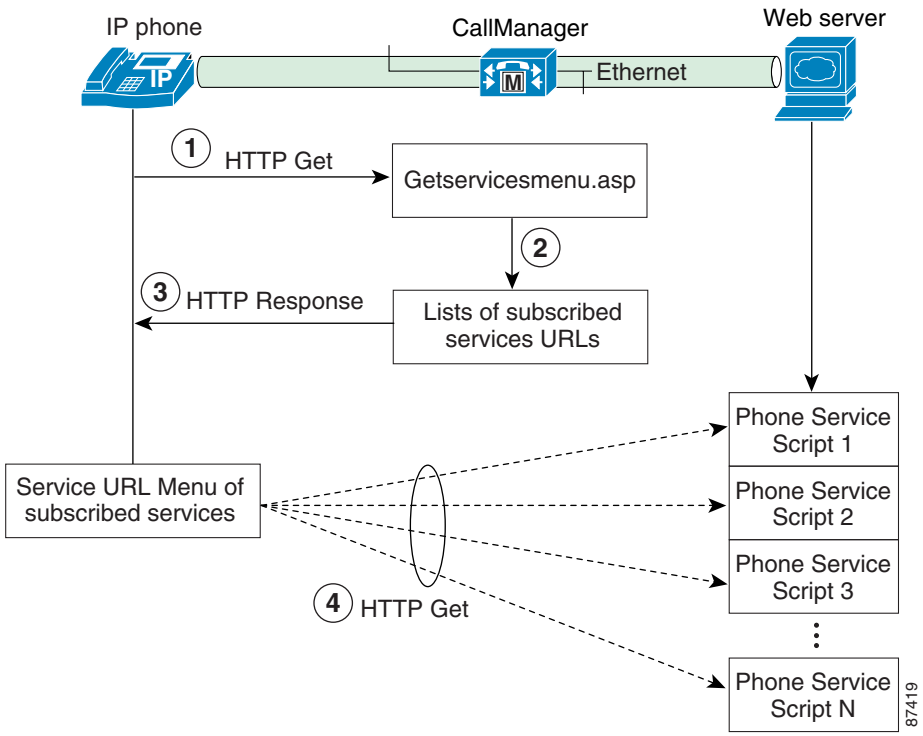
Description	Default Port Number or Range	Comments
HTTP proxy server requests	Port 80 for Microsoft IIS Port 8080 for Tomcat Servlet Engine	Default port varies with the web server platform
Access to Cisco CallManager	TCP port 8404	User-initiated press of the Services button to access the menu of user-subscribed phone services
Directory lookup	TCP port 389 for corporate LDAP directory TCP port 8404 for Cisco CallManager Domain Controller (DC) Directory	Port 389 is the default port for an LDAP-compliant directory server. Use TCP port 8404 if the phone service refers to the Cisco CallManager Domain Controller (DC) Directory.
IP Phone Service streaming media to/from the IP Phone	UDP port range: 20480 to 32768	These are the RTP port ranges already allocated by Cisco CallManager. By default, these ports are dynamically assigned, but the IP Phone service application can hard-code a designated RTP port. Only even ports in the range are allowed for use by phone service applications.

Redundancy

To ensure reliable services for phone users, you must maintain a high level of system availability, with a seamless transition to redundant systems during a system failure. While most of the back-end processing of a phone service occurs on a web server, the phones still depend upon Cisco CallManager to redirect them to the phone service.

Before discussing redundancy considerations, it is helpful to know how phone services are provisioned by default. As illustrated in Figure 11-3, when a user presses the Services button, an HTTP GET message is sent from the IP Phone to the Cisco CallManager getservicesmenu.asp script by default. (You can specify a different script by changing the setting in the Cisco CallManager enterprise parameters.) The getservicesmenu.asp script returns the list of phone service URL locations that are subscribed by the individual user. The HTTP response returns this list to the IP Phone. Any further phone service menu options chosen by the user continue the HTTP messaging between the user and the web server containing the selected phone service application.

Figure 11-3 Phone Services Managed by Cisco CallManager



Given the message flow in Figure 11-3, there are two main failure scenarios:

- Failure Scenario 1: Connection Fails to Cisco CallManager Server, page 11-6
- Failure Scenario 2: Connection Fails to Web Server Hosting the IP Phone Service, page 11-6

Table 11-3 summarizes these failure scenarios.

Table 11-3 Failure Scenarios for IP Phone Services

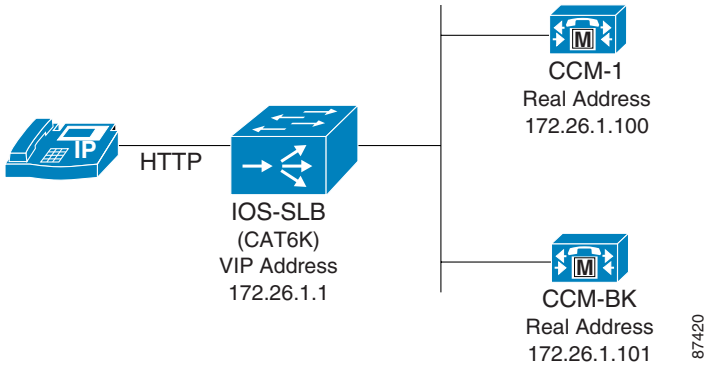
Failure Scenario	Behavior	Comments
Cisco CallManager fails	Failover to backup Cisco CallManager	Default behavior — There is no redundancy for user-initiated IP Phone services because the Services button is statically configured to a predetermined Cisco CallManager. Redundancy is available for Idle URL. Idle URL location and timeout value are saved on the TFTP server and downloaded to the IP Phone after the phone re-registers. Extension Mobility is preserved because the phone service device profile is preserved. Recommended implementation — Implement Server Load Balancing (SLB), either with Cisco IOS or with Cisco LocalDirector.
Web server fails	Failover to backup web server	Achievable through Cisco Server Load Balancing (SLB)

Failure Scenario 1: Connection Fails to Cisco CallManager Server

Redundancy in this case depends upon the functionality of the phone service as well as on how it is configured. When a primary Cisco CallManager server fails, all of the phones re-home to the backup Cisco CallManager server. However, the IP Phone configuration file downloaded from the TFTP server retains the same URL settings from the primary Cisco CallManager server. Therefore, the URL settings not dependent upon Cisco CallManager (such as URL Idle and URL Proxy) remain operational.

You can also provide redundancy by implementing server load balancing, either through Cisco LocalDirector or Cisco IOS Server Load Balancing (SLB) on a Catalyst 6000. (See Figure 11-4.)

Figure 11-4 Methods of Providing Redundancy for Phone Services



Failure Scenario 2: Connection Fails to Web Server Hosting the IP Phone Service

In this scenario, the connection to the Cisco CallManager server is preserved, but the link fails to the web server hosting the user-subscribed phone service. This is an easier scenario to provision for redundancy because the IP Phone is still able to access the Cisco CallManager server when the Services button is pressed. In this case, the IP Phone is similar to any other HTTP client accessing a web server. As a result, you can use any content distribution network (CDN) technology such as Server Load Balancing (SLB) to redirect the HTTP request from the phone to a hot standby web server.

Quality of Service

IP packets from the IP Phone must be marked appropriately in order to have its traffic classified. For priority queuing, Cisco recommends that the packets contain DSCP markings of AF31 and EF. (See Table 11-4.)

Table 11-4 Traffic Classification for IP Phone Services

Protocol	Port	DSCP Marking	Comments
SCCP	TCP 2000	AF31, EF	Call control
RTP	UDP 20480-32768	AF31, EF	Automatically set if Cisco CallManager initiates RTP streaming between two IP Phones. Not set if RTP streaming is initiated by an IP Phone service; therefore, traffic must be marked by the application.
HTTP	TCP 80	None by default	Traffic must be marked by the application.

CHAPTER 12

Computer Telephony Integration (CTI)

Computer Telephony Integration (CTI) applications include any IP telephony applications that use the Cisco CallManager CTI interface. These applications are generally written using either Telephony Application Programming Interface (TAPI) or Java Telephony Application Programming Interface (JTAPI). Some Cisco applications such as Attendant Console (AC) bypass these higher-level application program interfaces (APIs) and write directly to the CTI layer. The guidelines in this section apply to the common CTI applications such as Cisco IP SoftPhone Release 1.3, Cisco Personal Assistant, and the Cisco Customer Response Solutions (CRS) platform used for Cisco IP Interactive Voice Response (IVR) and IP Integrated Contact Distribution (ICD).

CTI applications are able to control the following Cisco CallManager device types:

- CTI ports
- CTI route points
- IP Phones

Scalability Guidelines

Use the following procedure to scale CTI resources for your applications.

Step 1 Identify the CTI devices required to run a specific CTI application.

Step 2 Calculate the average device weight for each CTI application. Table 12-1 shows the base device weights for the various CTI resources used by CTI applications.

Table 12-1 Base Device Weights for CTI Devices

Device	Base Device Weight	Comments
CTI Port	2	Base weight is for each CTI port line appearance.
IP phone (monitored and controlled as a third-party controlled phone, or 3PC)	3 for each controlled IP phone	Base weight is for each controlled IP phone. The value already includes the device weight of the IP phone.
CTI Route Point	2	

Step 3 When calculating device weights for CTI applications, estimate the busy hour call attempts (BHCA) for each line appearance to determine the BHCA factor. Increase the BHCA factor by 1 for every 6 BHCA. If the CTI application is client-based (for example, Cisco IP SoftPhone), 6 BHCA is a good default value.

Step 4 Provision devices across the servers in a cluster, according to the following guidelines:

- Maximum of 2500 CTI devices per Cisco CallManager server
- Maximum of 10,000 CTI devices per Cisco CallManager cluster

The following assumptions also apply to these guidelines:

- The cluster consists of Cisco MCS 7845 servers.
- Each CTI device is processing 6 or fewer BHCA.
- No other CTI applications requiring CTI devices are provisioned for the entire cluster, including CTI-dependent Cisco CallManager services such as automated alternate routing (AAR) and Cisco IP Manager Assistant (IPMA).

Redundancy

Table 12-2 summarizes the general behavior of CTI applications in various failover scenarios. Properly written CTI applications that handle the expected failure events should account for recovery, depending upon the software implementation of the application. Therefore, review the product documentation for the CTI application to be certain about its recovery capabilities.

Table 12-2 CTI Application Failover Scenarios

Failure Scenario	Behavior	Comments
Cisco CallManager fails	Failover to backup Cisco CallManager	CTI devices can be combined in device pools and assigned server priority within Cisco CallManager groups.
CTI Manager fails	Failover to backup CTI Manager	The primary and backup CTI Managers are assigned at each Cisco IP SoftPhone TAPI Service Provider (TSP) client configuration.
Cisco CallManager publisher fails	No redundancy if Domain Controller (DC) Directory is configured as the Cisco CallManager LDAP directory server.	Integration with a corporate directory server with a high-availability deployment will avoid this problem.
The CTI client or server application fails	Depends. Application availability and recovery is specific to the product's software implementation.	Refer to the specific product documentation for the CTI application.

Delay Considerations

Table 12-3 lists the average packet delays of common CTI activities.

Table 12-3 Average Packet Delays of Common CTI Activities

CTI Activity	Average Delay[1] (Seconds)	Comments
CTI initialization time	0.170	Values based on the average of multiple test results.
Time to dial tone	0.125	Values based on the average of multiple test results.
Time to establish Real-Time Transport Protocol (RTP) streams	0.308	Values based on the average of multiple test results.

1. Values are based on CTI baseline tests for Cisco CallManager Release 3.2.

Note The values in Table 12-3 should be used as an estimate only. The actual packet delays can vary, based on how a particular CTI application was developed. Additional back-end processing from a particular application can increase these delay times.

Quality of Service (QoS)

Table 12-4 summarizes the default traffic markings used by Cisco CallManager for CTI-based applications.

Table 12-4 Default Traffic Classifications for CTI-Based Applications

CTI-Related Messages	IP Precedence Value	Comments
CTI quick buffer encoding (QBE)	3	CTI QBE call control messages are the same for both TAPI and JTAPI-based applications.
Skinny Client Control Protocol (SCCP)	3	CTI QBE issues requests via the CTI Manager process to the Cisco CallManager (ccm.exe) process. Cisco CallManager then issues the call processing messages using SCCP.
Real-Time Transport Protocol (RTP)	5	RTP handles media streaming and control for the application.

Note The IP Precedence values in Table 12-4 are set if the request is handled by Cisco CallManager, but the traffic might not be classified by the application itself. Applications are responsible for marking IP Precedence for messages they are requesting. Refer to the product documentation for each CTI supporting application to ensure that the application is marking traffic appropriately.

CHAPTER 13

Cisco IP Interactive Voice Response (IVR)

This chapter summarizes the system design considerations for integrating Cisco IP Interactive Voice Response (IP IVR) with Cisco CallManager. IP IVR is based on the Cisco Customer Response Solutions (CRS) platform. The CRS platform contains multiple subsystems supporting protocol interfaces that allow application developers to build extensible and flexible voice and data services. Interfaces include JDBC, LDAP, HTTP, XML, ASR/TTS, and Cisco CallManager CTI (via the CRS JTAPI subsystem).

Note Unless stated otherwise, the information in this section applies to IP IVR as it is bundled with CRS Release 2.2.

Scalability

Cisco IP IVR has the following scalability considerations:

Call Sizing

Call sizing for IP IVR involves calculating how many IP IVR ports are needs to handle a certain number of calls. For proper sizing of IP IVR resources, consult with your Cisco account team.

CRS Server Scalability

You must size the CRS server to handle the performance impact on the CRS subsystems due to the CRS scripted application. For proper sizing of CRS server resources, consult with your Cisco account team.

Cisco CallManager Scalability

Once you determine how many IP IVR sessions you want to handle for calls, then determine how many Cisco CallManager resources are required for the CRS application. CRS applications may access one or more Cisco CallManager application interfaces. These interfaces can include authentication with Cisco CallManager's Domain Controller (DC) Directory using LDAP or with the Cisco CallManager database for retrieving or modifying device information using the AVVID XML Layer (AXL) API. These considerations are not covered in this document, but the impact varies depending upon how the CRS application was scripted.

Use the following procedure to estimate the Cisco CallManager capacity required for call processing from CTI resources used by IP IVR applications.

Step 1 Determine which of the following types of CTI resources are required for the IP IVR application:

- A *CTI route point* assigned to a main number (JTAPI trigger) for the IP IVR application
- A *CTI port* assigned to each session of the IP IVR application

Step 2 Estimate the average Busy Hour Call Attempts (BHCA) required for the IP IVR application.

Step 3 Calculate the average device weight for each IP IVR application, according to the following formula:

IP IVR Application Device Weight = (Base Device Weight) * (BHCA Factor) * (Number of devices)

Base Device Weight = 2 for each CTI route point or CTI port

BHCA Factor = (Estimated BHCA) / 6

Step 4 Provision devices across the servers in a cluster, according to the following guidelines:

- Maximum of 2500 CTI devices per Cisco CallManager server
- Maximum of 10,000 CTI devices per Cisco CallManager cluster

The following assumptions also apply to these guidelines:

- The cluster consists of Cisco MCS 7845 servers.
- Each CTI device is processing 6 or fewer BHCA.
- No other CTI applications requiring CTI devices are provisioned for the entire cluster, including CTI-dependent Cisco CallManager services such as automated alternate routing (AAR) and Cisco IP Manager Assistant (IPMA).

Redundancy

Table 13-1 summarizes IP-IVR behavior during various failure scenarios.

Table 13-1 Failure Scenarios for IP IVR

Failure Scenario	Behavior	Comments
Cisco CallManager fails	Failover to backup Cisco CallManager	IP IVR sessions and route points are combined in device pools and assigned server priority in Cisco CallManager group settings. It takes approximately 170 ms for each CTI device to re-register to the backup Cisco CallManager server after the heartbeat timeouts have expired and failure of the primary server has been detected.
CTI Manager fails	Failover to backup CTI Manager	It takes approximately 170 ms for each CTI device to re-register to the backup server after the heartbeat timeouts have expired and failure of the primary server has been detected.
Cisco CallManager publisher fails	No redundancy if Domain Controller (DC) Directory is configured as the Cisco CallManager LDAP directory server	Integration with a corporate directory server with a high-availability deployment will avoid this problem.
CRS server fails	No hot standby redundancy	A cold-standby solution of hardware swapping is available. Consult your Cisco account team for details.

Bandwidth Provisioning

CRS Release 2.2 supports only G.711 codecs. CRS Release 3.0 and later support both G.711 and G.729a codecs.

Cisco recommends G.729a low-bandwidth codec configurations for deployments where a significant number of calls originate from remote branches over an IP WAN connection that has limited bandwidth.

Quality of Service (QoS)

For IP IVR applications, the CRS software tags the following traffic:

- Call control traffic with CTI quick buffer encoding (QBE) — Type of Service (ToS) 3
- Real-Time Transport Protocol (RTP) traffic — ToS 5

CRS marks traffic for only the JTAPI subsystem. Other CRS subsystems that may be scripted for supporting IP IVR operations (for example, database queries, LDAP authentication, and CRS session ID traffic) are not classified and therefore are considered "best effort". For further details on the other CRS subsystems, refer to the *Cisco Customer Response Applications Developer's Guide*, available at

http://www.cisco.com/univercd/cc/td/doc/product/voice/sw_ap_to/apps_22/dvlpr/index.htm

Cisco IP SoftPhone

This chapter summarizes the following design considerations that apply when using Cisco IP SoftPhone with Cisco CallManager:

The information in this section applies explicitly to Cisco IP SoftPhone Release 1.3. For specific details about Cisco IP SoftPhone configuration and features, refer to the *Cisco IP Softphone Administrator Guide (1.3)*, available online at

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_ipphon/ip_7960/softphon/ver_1_3/english/index.htm

Scalability Guidelines

Cisco IP SoftPhone is a client desktop or laptop application that allows you to receive your office calls directly from your client machine.

Use the following procedure to scale Computer Telephony Integration (CTI) resources for Cisco IP SoftPhone.

Step 1 Determine what types of CTI resources are required for the Cisco IP SoftPhone. You can provision the Cisco IP SoftPhone in either of two ways:

- By assigning a CTI port to the Cisco IP SoftPhone line appearance
- By using the Cisco IP SoftPhone line appearance to control a hardware IP phone

Step 2 Calculate the average device weight for each Cisco IP SoftPhone user. Table 14-1 shows the base device weights for the Cisco IP SoftPhone CTI resources.

When calculating the Cisco IP SoftPhone device weights, follow these additional guidelines:

- Estimate the busy hour call attempts (BHCA) for each line appearance to determine the BHCA factor. Increase the BHCA factor by 1 for every 6 BHCA. For general use, 6 BHCA is a good default value.

- Consider the call handling for the majority of the Cisco IP SoftPhone clients. If most of the calls handled per Cisco IP SoftPhone session involve a transfer or conferencing, then you must factor in a call handling multiplier. If not, then the call handling multiplier is negligible and can be assigned a value of 1.

Table 14-1 Cisco IP SoftPhone Base Device Weight Values

Cisco IP SoftPhone Configuration	Base Device Weight	Comments
Standalone Cisco IP SoftPhone (no associated hardware IP Phone)	2 for each line appearance	Base weight is for each CTI port line appearance.
Cisco IP SoftPhone controlling a hardware IP Phone	3 for each controlled hardware IP Phone	Base weight is for each controlled hardware IP Phone. The value already includes the device weight of the IP Phone.

Step 3 Provision the devices across the servers in the cluster, according to the following guidelines:

- Maximum of 2500 Cisco IP SoftPhones per Cisco CallManager server (requires MCS 7845 servers for the Cisco CallManager cluster)
- Maximum of 10,000 Cisco IP SoftPhones per single Cisco CallManager cluster (requires MCS 7845 servers for the Cisco CallManager cluster)

The following assumptions also apply to these guidelines:

- Each Cisco IP SoftPhone is configured with only 1 line appearance.
- Each Cisco IP SoftPhone is processing no more than 6 BHCA.
- No other CTI applications requiring CTI devices are provisioned for the entire cluster. This assumption includes CTI-dependent Cisco CallManager services such as automated alternate routing (AAR) and Cisco IP Manager Assistant (IPMA)

Redundancy

Redundancy for the Cisco IP SoftPhone is similar to that of a hardware IP phone, except that the Cisco IP SoftPhone interfaces primarily with the CTI Manager for its failover handling.

Table 14-2 summarizes the anticipated Cisco IP SoftPhone behavior in various failover scenarios.

Table 14-2 Cisco IP SoftPhone Failover Scenarios

Failover Scenario	Behavior	Notes and Comments
Cisco CallManager fails	Failover to backup Cisco CallManager	Cisco IP SoftPhones are combined in device pools and assigned server priority in Cisco CallManager group settings.
CTI Manager fails	Failover to backup CTI Manager	The primary and backup CTI Managers are assigned at each Cisco IP SoftPhone TAPI Service Provider (TSP) client configuration.
Cisco CallManager publisher server fails	No redundancy if Domain Controller (DC) Directory is configured as the Cisco CallManager LDAP directory server	Integration with a corporate directory server with a high-availability deployment will avoid this problem.
Cisco IP SoftPhone fails	No redundancy to another Cisco IP SoftPhone on the same PC client	For a standalone Cisco IP SoftPhone configuration with a CTI port, calls in progress will be lost. New incoming calls may be preserved by configuring the Call Forward No Answer (CFNA) number to another IP phone extension. For a Cisco IP SoftPhone controlling a hardware IP phone, calls in progress will be preserved to the hardware IP phone.

Bandwidth Provisioning

Cisco IP SoftPhone supports G.711 and G.729a codecs.

Cisco recommends G.729a low-bandwidth codec configurations in the following cases:

- Deployments with telecommuters connecting their Cisco IP SoftPhones over a Virtual Private Network (VPN).
- A centralized call processing deployment model where there are Cisco IP SoftPhone users that "roam" among branch offices. It is also necessary in this case to over-provision bandwidth for these roaming users. (See Multi-Site WAN with Centralized Call Processing, page 1-4.)

Quality of Service

The current version of Cisco IP SoftPhone does not classify its call control traffic or it media traffic. However, Cisco CallManager sets default traffic classifications as indicated in Table 14-3.

Table 14-3 Default Traffic Classifications for CTI-Based Applications

CTI-Related Messages	IP Precedence Value	Comments
CTI quick buffer encoding (QBE)	3	CTI QBE call control messages are the same for both TAPI and JTAPI-based applications.
Skinny Client Control Protocol (SCCP)	3	CTI QBE issues requests via the CTI Manager process to the Cisco CallManager (ccm.exe) process. Cisco CallManager then issues the call processing messages using SCCP.

Security

This chapter is not a complete treatment of network security. Rather, it is a starting point for network managers, system administrators, and systems engineers to use when building IP telephony networks that adhere to the Cisco SAFE security model. For more information, refer to the *Cisco SAFE White Paper,* available at

http://www.cisco.com/warp/public/cc/so/cuso/epso/sqfr/safe_wp.htm

The following list summarizes the major security guidelines and recommendations described in this chapter:

Establish a Corporate Security Policy

A security policy can be as simple as an acceptable use policy for network resources or can be several hundred pages in length and detail every element of connectivity and associated policies. Although somewhat narrow in scope, Request For Comments (RFC) 2196 suitably defines a security policy as follows:

> *A security policy is a formal statement of the rules by which people who are given access to an organization's technology and information assets must abide.*

The following information can assist you in developing a corporate security policy:

- RFC 2196 *Site Security Handbook*, available at http://www.ietf.org/rfc/rfc2196.txt
- A sample security policy for the University of Illinois, available at http://www.aits.uillinois.edu/security/securestandards.html
- The SysAdmin, Audit, Network, Security (SANS) Institute Security Policy Project, available at http://www.sans.org/resources/policies/

It is important to understand that network security is an evolutionary process. No one product can make an organization secure. True network security comes from an array of products and services, combined with a comprehensive security policy and a commitment to adhere to that policy at all levels of the organization. In fact, a properly implemented security policy without dedicated security hardware can be more effective at mitigating the threat to enterprise resources than a comprehensive security product implementation without an associated policy.

It is also important to approach security in layers, such that the compromise of any one network or security technology does not result in the compromise of the entire solution. Thus, securing voice is not just about securing voice, just about securing the data network, just about installing a firewall, or just turning on user authentication. To properly secure voice, you must adopt a systems approach covering all elements of the voice and data network. The security policy must cover all associated technologies along with plans for monitoring, providing remedies, and social engineering.

Provide Physical Security

As with most computing devices, Cisco routers, switches, servers, and other infrastructure components are not designed to provide any protection against penetration or destruction by an attacker with direct physical access. You must take reasonable steps to prevent physical access by unauthorized personnel.

Strategic corporate assets such as the data network must be secured physically and adequately. Make sure that access to network equipment is limited, wiring closets are locked, and wires are not in plain view. Mission-critical resources might require geographical separation to provide effective spacial redundancy. Also remember to secure the power plant, because killing power can be one of the most effective forms of disrupting service.

Protect the Network Elements

Once you have established physical security, the next step is to secure the actual routers, switches, and Voice over IP (VoIP) gateways that make up the communications network. These network elements provide both physical and logical connectivity of the entire enterprise network and, according to the SAFE architecture, should be considered as a target for well-informed attackers. For recommendations and configuration details on securing these infrastructure devices, refer to *Improving Security on Cisco Routers*, available at

http://www.cisco.com/warp/customer/707/21.html

To protect the network elements, perform the following tasks:

- Secure Login Access, page 15-3
- Follow Sound Password and Authentication Practices, page 15-3
- Assign Unique Port VLAN ID (PVID) to Each 802.1Q Trunking Port, page 15-3
- Ensure That Unused Router Services Are Disabled, page 15-3
- Securely Configure Network Management Functions, page 15-4
- Use Logging Services to Track Access and Configuration Changes, page 15-4

Secure Login Access

You can perform configuration from the Command Line Interface (CLI) via telnet sessions, but Secure Shell (SSH) is the preferred method of managing routers and switches. The first step in securing the network elements is to limit which subnets are able to access the router and switch virtual terminal sessions. Limiting virtual console access to the IP address range(s) of operations staff and network management hosts is a useful method to restrict unauthorized users from accessing network devices, even if a password is discovered.

Follow Sound Password and Authentication Practices

Passwords are another important line of defense against unauthorized access to routers and switches. The best way to handle user passwords is to use a TACACS+ or RADIUS authentication server in conjunction with one-time password systems such as SofToken, SecureID, or DES Gold Cards to prevent an attacker from reusing trusted user passwords. However, many routers still have a locally configured password for privileged access during required maintenance periods.

To ensure maximum precautions for limiting router password exposure, use the **service password-encryption** command to encrypt all passwords. Cisco also recommends that you use the **enable secret** command to hide configuration access even further. For maximum security, use numbers and punctuation symbols as well as mixed-case letters in passwords. Finally, use an encrypted form of access, such as IPSec or Secure Shell (SSH), for administering network devices.

Assign Unique Port VLAN ID (PVID) to Each 802.1Q Trunking Port

An often overlooked aspect of designing campus networks is the securing of 802.1Q Ethernet trunks. A trunk is an Ethernet connection between two network devices that carries multiple virtual LANs (VLANs). Potential security threats on 802.1Q VLAN trunk configurations were brought to light in a September 1999 BUGTRAQ alert (BUGTRAQ 801.1Q Security Alert; 9/99 BUGTRAQ@securityfocus.com email; Subject: *VLAN Security*). This alert pointed out that, if a user's native VLAN ID is the same as the port VLAN ID (PVID) of the 802.1Q trunk, then the user can send frames from his VLAN and have them "hop" to other VLANs. This weakness is part of the 802.1Q specification and does not apply to Cisco ISL trunking ports.

The workaround for this threat is to ensure that every 802.1Q trunking port has a PVID, or native VLAN ID, that is unique throughout the campus network.

Ensure That Unused Router Services Are Disabled

Many of the unnecessary services that can run on Cisco routers are disabled (turned off) by default in Cisco IOS Release 12.1 and later. However, it is always a good practice to audit the network and ensure that these services are disabled. Disable the following services if they are not used:

- Hypertext Transfer Protocol (HTTP)
- Finger
- User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) Small Server
- Remote Copy Protocol (RCP) or Remote Shell Protocol (RSH)
- Telnet

Securely Configure Network Management Functions

Simple Network Management Protocol (SNMP) is widely used for monitoring and configuring network elements. SNMP uses an authentication method based on a *community string*. This community string is essentially a password used for accessing the network element. Since nearly all of the information viewable or configurable from a virtual console can also be accessed with SNMP, it is essential to restrict this method of access as completely as possible.

The following basic rules improve SNMP security:

- Never use *public* and *private* as community strings.
- Limit SNMP access to only a few specific hosts or subnets.

Use Logging Services to Track Access and Configuration Changes

Accurate logging is still one of the most valuable tools for tracking intruders. Logging all system notices and error messages often provides valuable insight into the operational status of network devices. If you log access list violations, you can also correlate the logs between devices to determine if the network is being probed or if a device has been compromised. To obtain an accurate log, perform the following steps on all network elements:

- Configure all devices to use an accurate, centralized time source, such as an authenticated NTP server.
- Enable time-stamping on all Cisco IOS and CatOS devices.
- Designate a syslog server to receive logging information.

Design a Secure IP Network

For an IP telephony system to be secure, it has to be built upon a secure data network. Before installing IP phones on the network, spend time on designing a sound and secure IP network. Think about using separate broadcast domains, building logical associations of IP telephony equipment, isolating the IP telephony management servers, establishing security relationships, and building perimeters secure from both outside attackers and internal users. When building a secure IP telephony network, follow these guidelines:

- Place all Cisco CallManagers, IP telephony application servers, and IP telephones on their own, separate IP networks.
- Ideally, these subnets should use a different major address range than the corporate data networks. Where possible, use RFC 1918 IP address space, which cannot be routed to the Internet, to further separate the IP telephony networks. Use Network Address Translation (NAT) judiciously to provide translation between the voice and data IP networks, and configure it only where required by Call Center applications, Cisco IP SoftPhone, or Cisco WebAttendant. The Internet gateway router or firewall should not allow NAT translations for Internet-to-IP telephony connectivity.
- Protect voice at Layer 2.
- Use access control lists (ACLs) on the gateway router between the IP telephony network and the enterprise data network to eliminate any well-known, malicious attacks that might originate from within the corporate network.
- Place firewalls in front of the Cisco CallManager clusters.
- Use application layer gateways (ALGs) for stateful inspection of voice through firewalls and NATs.

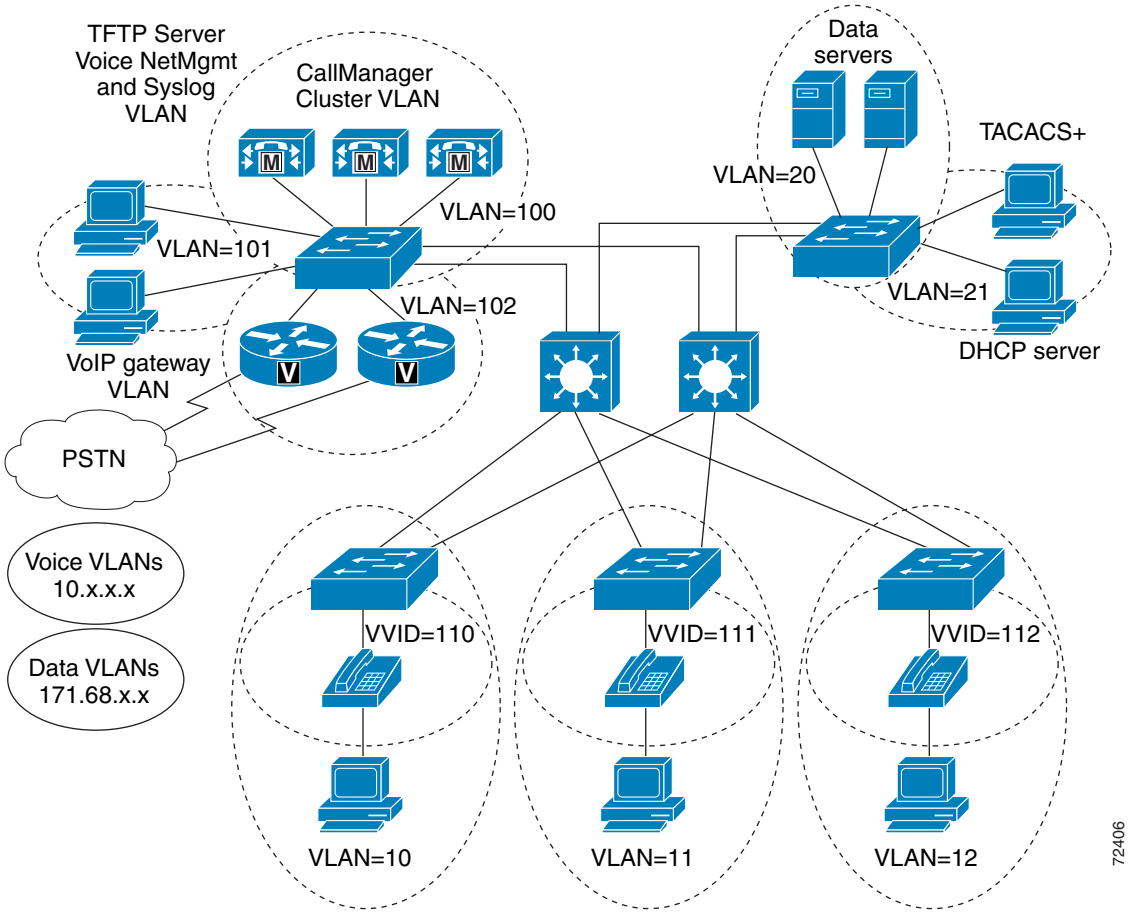
The following sections provide more information on designing a secure IP network:

Creating and Assigning VLANs and Broadcast Domains

Many IP security solutions can be implemented only if a packet encounters a Layer 3 (IP) device. The Cisco CallManager cluster, IP telephones, VoIP gateways, and network management workstations should always be on their own subnets, separate from the rest of the data network and from each other. In fact, every device should use a separate segment to connect to the network. Using separate segments for devices (in other words, a switched Ethernet infrastructure) prevents any attacker or attacking application from snooping or capturing other Ethernet traffic as it traverses the physical wire. In addition, the recommended Cisco AVVID design model uses separate subnets for an IP phone and its attached data PC by using 802.1Q VLAN trunking technology.

Figure 15-1 depicts the major components in a typical enterprise network. In this example, all IP telephony components reside on various subnets and VLANs in the voice IP network (10.x.x.x), and all data pieces such as PCs, email servers, and the DHCP server, reside on the data IP network (171.68.x.x). In addition, this network is a 100% switched Ethernet environment with every user and device residing on a separate segment.

Figure 15-1 Typical Enterprise Network



Protecting Voice at Layer 2

This chapter does not go into the details below Layer 2 security, but it attempts to make you aware of a number of Layer 2 technologies for securing DHCP, ARP, and other protocols at Layer 2. Some of the available technologies for securing Layer 2 include Private VLANs, Port Security, DHCP Snooping, IP Source Guard, Secure ARP Detection, and Dynamic ARP Inspection.

Note Private VLANs and Port Security do not support trunk ports and therefore cannot be used on ports with auxiliary VLANs configured for phones. They can, however, be used effectively on ports connected to Cisco CallManager, Cisco Unity, and other servers.

For details on securing Layer 2, refer to the *SAFE Blueprint*, available at

www.cisco.com/go/safe

Implementing Packet Filters

Using IP filters has been an integral piece in building secure networks for several years. Cisco highly recommends that you use filters when securing IP telephony networks in order to limit access to the voice network. In most enterprises, you would place these filters on the router connecting the IP telephony and data networks. Filters can help with the following types of security issues:

Directed Broadcasts

If a Cisco interface is configured with the **no ip directed-broadcast** command, directed broadcasts that would otherwise expand into link-layer broadcasts at that interface are dropped instead.

Source-Routed Packets

A Cisco router with **no ip source-route** set will never forward an IP packet that carries a source routing option. Use this command on the router connected to the Internet as well as on the gateway router connecting the IP telephony and data networks within the enterprise.

ICMP Redirects

Cisco recommends using an access list on the border router between the IP telephony network and the data network to filter all Internet Control Message Protocol (ICMP) redirects.

TCP Intercept

To help protect the VoIP network from Denial of Service (DoS) attacks, configure an access list on the gateway router between the IP telephony and data networks to match any destination IP addresses in the voice network. Then apply the **ip tcp intercept list** command to the Ethernet interface to look for suspicious TCP SYN traffic.

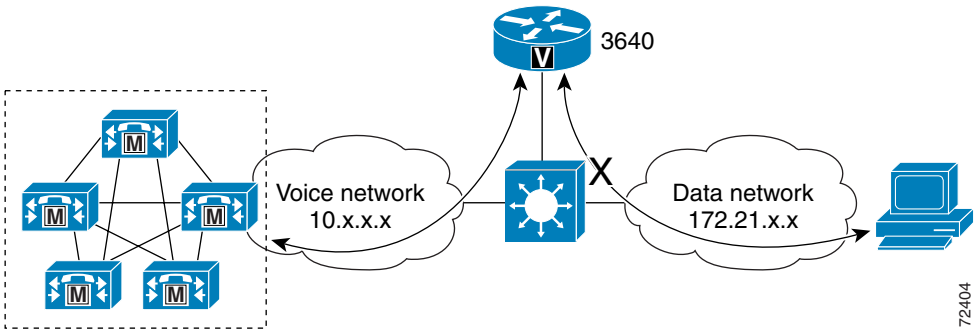
Reverse Path Forwarding (RPF)

By discarding IP packets that lack a verifiable IP source address, the Unicast RPF feature helps mitigate problems caused by the introduction of malformed or forged (spoofed) IP source addresses into a network. For descriptions of other commands used when configuring Unicast RPF, refer to the *Cisco IOS Command Reference Master Index* or search online at http://www.cisco.com.

Protecting the VoIP Gateways

As illustrated in Figure 15-2, it is important to allow call signaling only from Cisco CallManager (or, when using toll bypass, from other VoIP gateways). The easiest way to block direct call setup attempts from other devices is through the use of access control lists (ACLs) on either an upstream router or the VoIP gateway itself, thus limiting incoming signaling requests to those calls that originate on Cisco CallManager.

Figure 15-2 Preventing Direct Calls from the Data Network



Permitting Other Services

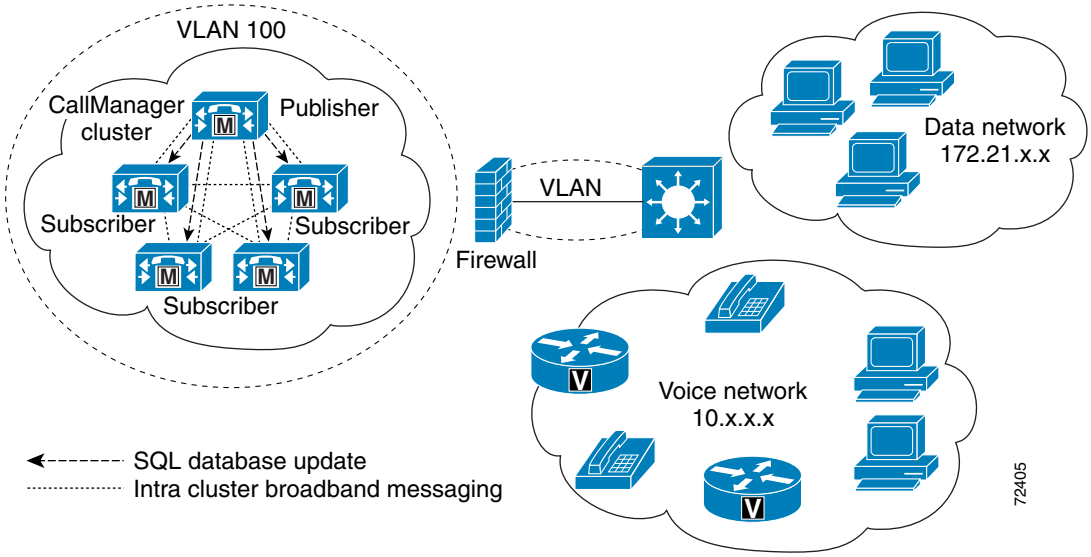
Many enterprises require at least minimal communications between the data network and the IP telephony network for policy, application, and cost reasons. Cisco is currently updating the documentation of UDP and TCP port utilization for each of our voice products. This new documentation includes more details to allow stricter access control lists (ACLs) to be deployed in customer networks.

At the time of this publication, the new port utilization documentation was not complete. To obtain this information, contact your Cisco Systems Engineer (SE) or account team.

Firewalls

Internet firewalls are a default piece of network infrastructure. This document assumes that you have already implemented an Internet security policy and architecture using network design, firewalls, and intrusion detection applications. Sound security policies dictate that any partner connections require additional firewall measures. Once these basic firewalls are in place, and you have built the AVVID network and connected it to the existing IP network, you should add another firewall between the Cisco CallManager cluster and the IP telephony and data networks, as illustrated in Figure 15-3.

Figure 15-3 Building a Firewall Around Cisco CallManager



Note Cisco does not recommend placing a firewall between Cisco CallManager servers within the same cluster, such as in a deployment using clustering over the WAN.

By placing a firewall between the Cisco CallManager cluster and both the voice and data networks, you greatly reduce the exposure of the most critical component in the Cisco AVVID network, the call processing agent. The firewall acts as a guardian between all IP devices and the Cisco CallManagers, ensuring that only specific transactions are allowed.

Note If you are using H.225 gatekeeper-controlled trunks, calls across the firewall might experience potential problems with one-way audio and/or call setup failure. To avoid these problems, configure the Cisco CallManager Service Parameter for the H.225 gatekeeper-controlled trunks to use port 1720 to register with the gatekeeper. Also configure the firewall to permit any TCP packets from outside the firewall to be received by the Cisco CallManager inside the firewall on port 1720.

For more information on firewalls, refer to the Cisco PIX documentation, available online at

http://www.cisco.com/univercd/cc/td/doc/product/iaabu/pix/index.htm

Also refer to the Cisco IOS Firewall documentation at

http://www.cisco.com/warp/public/cc/pd/iosw/ioft/iofwft/

Application Layer Gateway (ALG)

An application layer gateway (ALG), sometimes called an inspection engine or fixup) performs the stateful inspection of certain applications to facilitate their traversal of firewalls and Network Address Translations (NATs). ALGs exist for many different types of applications. For example, in the case of voice, there are ALGs for SIP, SCCP, MGCP, and H.323. These ALGs enable voice packets to traverse the firewalls and NATs successfully.

Secure Cisco CallManager

Cisco CallManager runs on a server that uses the Microsoft Windows 2000 Server operating system. All OS hardening that is required is applied during the installation procedure. All services that are running after the install are required for the system to operate.

The following sections describe some of these services in more detail:

Securing Windows

A great deal has been already done to secure the Microsoft Windows operating system that comes with Cisco CallManager. The registry, NTFS file system, IIS service, and many other aspects of the operating system are locked down, and no further security configuration is needed or supported. Any other services that you start will impact the security of the system. Configure only the software that is approved to be installed on the Cisco CallManager servers. If you install any other applications on the Cisco CallManager servers, they could affect the performance and security and would have to be uninstalled before Cisco support services could help in troubleshooting if there are any system problems.

Disable Unused Windows Services

Most services and security are configured to run during the install, and you should leave them as installed. However, there are a few additional services that you can disabled (turned off) to improve the security of Cisco CallManager. Most computer attacks are against the IIS service. Because Cisco recommends that all administration be done on the publisher server, you can disable IIS on all of the subscribers in a Cisco CallManager cluster. Phones register with the subscribers, so by disabling IIS on those servers, you greatly diminished access from malicious attacks to the servers that would have the greatest production impact on your system. IIS is needed during software upgrades, so the administrator must remember to re-enable it when upgrading. If IIS is set to manual, the upgrade procedure can enable IIS when it is needed. In addition to IIS, DHCP and TFTP can also be disabled on all of the servers except where they are specifically needed, usually on the publisher.

User Accounts and Passwords

The administrator account on Cisco CallManager is a Windows NT account stored in the Microsoft Security Account Manager (SAM) database, and it uses the same authentication mechanisms as Windows. All usernames and passwords defined in Cisco CallManager Administration, such as those used for CTI or Multilevel Administration Access (MLA), are stored in the Domain Controller (DC) Directory. All passwords should be a minimum of six characters and use a non-pronounceable or non-predictable combination of letters and numbers.

Secure Administration

In addition to the Internet and IIS, Cisco supports the use of Microsoft Terminal Services or Virtual Network Computing (VNC) for remote administration. VNC is a client/server software package allowing remote network access to graphical desktops. With VNC, you can access your machine from anywhere, provided that your machine is connected to the Internet. VNC is free and is available on most platforms. The University of Cambridge Engineering Department lists configuration guidelines for securing VNC using Secure Shell (SSH), on their website at

http://www.uk.research.att.com/vnc/sshvnc.html

Keep Operating System Patches Up-to-Date

Unfortunately, viruses, worms, and denial-of-service attacks are now a part of daily life with computers. We often hear of, and experience, the proliferation of Smurf, Code Red, Nimbda, SQL Slammer, Blaster, Nachi, Sobig, or other viruses. Anti-virus and intrusion detection systems help to protect us against the atrocities of these attacks, but the best way to mitigate these attacks is to keep the operating system up-to-date.

Cisco monitors a variety of security alert services. Any new security vulnerability deemed critical or Severity 1 is fixed, tested, and posted to http://www.cisco.com within 24 hours. All other identified vulnerabilities that could impact Cisco CallManager are rolled into a monthly operating system patch.

The following applications all use the same operating system from Cisco:

- Cisco CallManager
- Cisco Conference Connection
- Cisco Emergency Responder
- Cisco IPCC Express
- Cisco IP IVR
- Cisco Internet Service Node (ISN)
- Cisco Personal Assistant

Operating system patches posted and announced by Cisco are valid for all of these applications. Apply patches to these applications only if the patches come from http://www.cisco.com.

Some websites direct customers of Cisco Unity and Cisco IPCC Enterprise (and other Cisco products not mentioned here) to download patches directly from the Microsoft website. Do *not* download the patches directly from Microsoft, but instead obtain the appropriate patches from the following Cisco website:

http://www.cisco.com/kobayashi/sw-center/sw-voice.shtml

Note Microsoft posts many patches that do not apply to Cisco CallManager, such as those for Windows 95/98 or Office suite products. Those patches are not incorporated into an operating system build.

For more information about Cisco's policy on security patches, refer to the *Cisco CallManager Security Patch Process* document, available at

http://www.cisco.com/application/pdf/en/us/guest/products/ps556/c1167/ccmigration_09186a0080157c73.pdf.

Cisco also has an email alias for notifying customers about the availability of new OS updates, with information about what is in the build, what attacks it mitigates, and why customers should apply it to their Cisco CallManagers. For information about how to subscribe to this alias, refer to

http://www.cisco.com/warp/public/779/largeent/software_patch.html

Cisco Emergency Responder and Cisco Conference Connection use the same operating system build as Cisco CallManager. As a result, all patches and upgrades to the Cisco CallManager OS may also be applied to these other two applications.

Virus Scanning on Cisco CallManager

Every Windows 2000 server, including Cisco CallManager, needs anti-virus software protection. At this time, Cisco does not bundle or include anti-virus software with Cisco CallManager. Through the AVVID Partner Program, Cisco supports the use of Symantec AntiVirus 7.6 or 8.0 and McAfee NetShield 7.0. You should configure the appropriate virus scanning software for the version of Cisco CallManager that you are running. For information on installing and configuring the virus scanning software, refer to the documentation available online at

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/sec_vir/index.htm

Cisco Security Agent Host-Based Intrusion Detection

The advanced Cisco Security Agent product provides threat protection for server and desktop computing systems. Cisco Security Agent goes beyond conventional host and desktop security solutions by identifying and preventing malicious behavior before it can occur, thereby removing potential known and unknown (day zero) security risks that threaten enterprise networks and applications. Cisco Security Agent aggregates and extends multiple endpoint security functions by providing host intrusion prevention, distributed firewall, malicious mobile code protection, operating system integrity assurance, and audit log consolidation, all within a single agent package.

As high-visibility network security attacks such Code Red and the SQL Slammer worm have shown, traditional host and desktop security technologies are limited in their capability to combat the effects of new and evolving attacks. Customers require robust endpoint security that prevents security attacks from affecting the network and critical applications. Because Cisco Security Agent analyzes behavior rather than relying on signature matching, its solution provides robust protection with reduced operating costs.

Beginning with Cisco CallManager Release 3.3(3), a non-managed version of Cisco Security Agent is bundled with Cisco CallManager at no additional charge. This version provides all the security benefits of fully-managed agents without the other features of those agents. Non-managed versions can be upgraded to fully-managed versions with the purchase of the Management Console, bundled with CiscoWorks VPN/Security Management Solution (VMS) Release 4.0, for a nominal charge. The Management Console, with fully-managed agents, centralizes the distribution and administration of

policies, provides reporting capabilities, and correlates events such as ping sweeps or port scans that occur on disparate systems. For more information about Cisco Security Agent, refer to the documentation at

http://www.cisco.com/en/US/partner/products/sw/secursw/ps5057/index.html

While Cisco Security Agent provides a strong measure of security against denial-of service, worms, viruses, or other illicit attacks, it does not represent a complete security solution. Effective protection for Cisco CallManager also depends on anti-virus software and on keeping the operating system patched and up-to-date.

Off-Load IP Phone Services

Most IP phone service XML applications go to the Internet to get up-to-date data for a variety of purposes. Because this Internet access can compromise security, you should remove all IP phone service applications from the publisher or subscribers in a Cisco CallManager cluster and put them on their own dedicated servers, either outside the corporate firewall or preferably in a neutral zone (DMZ).

Disable Auto-Registration of IP Phones

Auto-registration is a Cisco CallManager feature that enables previously uninstalled phones to boot, register automatically with Cisco CallManager, obtain a phone number, and become active on the network. This feature is a convenience, especially when you are installing large quantities of phones with the plug-in applet Tool for Auto-Registered Phone Service (TAPS). However, it also represents an opening for toll fraud, a malicious directed attack, or a denial-of-service attack. Cisco recommends that you disable auto-registration on all Cisco CallManager servers. You can use it temporarily to install large quantities of phones, but disable it again whenever the installation is complete.

Multi-Level Administration

Multilevel Administration Access (MLA) is a Cisco CallManager feature that provides role-based administration of Cisco CallManager. With MLA, you can create administrative groups, assign administrative users to those groups, and then grant privileges to the groups to deny access or to allow read/write or read-only access to individual Cisco CallManager Administration pages. For example, network administrators could have read/write access to all administration pages, but technicians could have read/write access only to the phone administration pages so that they can install phones, and operators could have read-only access to phone, gateway, and voice mail pages to check the current status of those devices.

Toll Fraud Prevention

Toll fraud is a serious issue in the telecommunications industry. The fraudulent use of telecommunications technology can be very expensive for a company, and it is essential that the telecom administrator take the necessary precautions to this fraud.

To help prevent toll fraud, configure your dial plan with the appropriate *calling search spaces* and *partitions* in Cisco CallManager. These constructs enable you to specify which dialed patterns can be reached by specific calling devices. An IP phone, for example, could have a calling search space that

includes a partition with a 9.xxxxxxx pattern in it. If this is the only pattern and the only partition associated with the calling search space on that phone, then only this pattern of digits can be dialed by that phone.

The basic measures for preventing toll fraud include:

- Using a dial plan to reduce or eliminate toll fraud
- Blocking the transfer of calls to extension 9xxx
- Disabling IP phone auto-registration
- Restricting the calling search spaces of call forward all, voice mail, and Personal Assistant
- Blocking area codes commonly used for toll fraud

For more information on designing a dial plan, refer to the chapter on Dial Plan, page 7-1. Also refer to the *Cisco CallManager Fundamentals* book, available from www.fatbrain.com or www.amazon.com.

Software MTP and Conferencing Services

Cisco recommends that you do not install the software-based media termination point (MTP) and conferencing services on the Cisco CallManager server. These applications terminate Real-Time Transport Protocol (RTP) and User Datagram Protocol (UDP) VoIP streams and mix them together to create another call leg or a conference call. The risk is that UDP is a difficult protocol to secure, and terminating it on the Cisco CallManager server exposes the server to attacks unnecessarily. To mitigate this risk, use either hardware-based MTP and conferencing or install the conferencing software on a separate Windows 2000 server.

System Auditing and Logging

Auditing lets you track the usage of many privileged tasks in Windows 2000. When auditing is enabled, regularly reviewing the Event Viewer can help you determine if the system has been compromised. Table 15-1 shows a suggested auditing scheme.

Table 15-1 Suggested Auditing Scheme

Description	Log Access	Log Failure
Audit Account Login Events	Yes	Yes
Audit Account Management	Yes	Yes
Audit Directory Service Access	Yes	Yes
Audit Login Events	Yes	Yes
Audit Object Events	No	Yes
Audit Policy Change	Yes	Yes
Audit Privilege Use	Yes	Yes
Audit Process Tracking	No	Yes
Audit System Events	Yes	Yes

Cisco CallManager SNMP

As documented above in Protect the Network Elements, page 15-2, SNMP community strings are passwords used to access a network element. Because nearly all of the information viewable or configurable from a virtual console can also be accessed with SNMP, it is essential to restrict this method of access as completely as possible.

The following basic rules improve SNMP security:

- Never use *public* and *private* as community strings.
- Limit SNMP access to only a few specific hosts or subnets.

Secure IP Phones

Use the following methods to protect IP phones from malicious attacks:

- Protect IP Phones from Gratuitous Address Resolution Protocol, page 15-15
- Isolate the Voice VLAN from the Attached PC, page 15-15
- Prevent Access to Network Configuration Information, page 15-16
- Disable the PC Port if It is Not Needed, page 15-16
- Ensure that the IP Phone Firmware is Valid, page 15-16

In addition, beginning with Cisco CallManager Release 3.3(3), all phone firmware images carry a digital signature for authenticity.

Protect IP Phones from Gratuitous Address Resolution Protocol

By default, IP phones accept Gratuitous Address Resolution Protocol (GARP) packets. Some devices use GARP to announce their presence on the network. However, attackers can also use GARP to spoof a valid network device. For instance, an attacker could send out a GARP message claiming to be the default router. Cisco CallManager Release 3.3(3) allows an administrator to disable the acceptance of GARP packets on the Phone Configuration page.

Note A phone learns the IP addresses of the devices it needs to know about in the DHCP load process. It then uses Address Resolution Protocol (ARP) to determine the MAC addresses of the devices with which it needs to communicate.

Disabling the acceptance of GARP specifically breaks the attack tool Ettercap.

Isolate the Voice VLAN from the Attached PC

By default, IP phones pass *all* packets received on the switch port (the one facing the upstream switch) to the PC port, including 802.1Q tagged packets destined for the phone. Cisco CallManager Release 3.3(3) allows an administrator to disable the forwarding of packets tagged with the voice VLAN to the PC port. Likewise, packets received from the PC port that are tagged with the voice VLAN will be dropped. This option allows a device attached to the PC port to use 802.1Q (if available or desired) without having access to the voice VLAN. You can configure this option on the Phone Configuration page. Disabling access to the voice VLAN from the PC port specifically breaks the attack tool VOMIT.

Note The Cisco IP Phone 7912 does not support this feature to disable the forwarding of voice packets to the PC port.

Prevent Access to Network Configuration Information

By default, pressing the Settings button on an IP phone provides access to configuration elements on the phone. Cisco CallManager Release 3.3(3) allows an administrator to disable the Settings button, thus prohibiting access to configuration information. You can configure this option on the Phone Configuration page.

Disable the PC Port if It is Not Needed

By default, the PC port is enabled on all IP phones that have a PC port. Cisco CallManager Release 3.3(3) allows the administrator to disable the PC port. This option is useful for lobby or conference room phones. You can configure this option on the Phone Configuration page.

Ensure that the IP Phone Firmware is Valid

The IP phone contains a feature that enables it to validate that the image it receives is an image generated by Cisco. This feature helps prevent Trojan-horse images from being installed in phones and subverting other protection mechanisms. Images are digitally signed and wrapped into a Digital Signature Envelope. When an IP phone downloads a new image, it compares the signature of that image with the signature in its existing image. If they do not match, the new image is rejected. Beginning with Cisco CallManager Release 3.3(3), all images are signed by default with no configurable options.

Note Once a signed image has been loaded into a phone, it cannot be removed or backed out.

The phone firmware images that support all of these features will also work with Cisco CallManager Release 3.3(2), but none of the configurable options will work, and their default behavior will be exactly like the firmware for Cisco CallManager Release 3.3(2); that is, GARP packets will be accepted by the phone, packets tagged with the voice VLAN will be passed to and accepted from the PC port, and both the PC port and the Settings button will be enabled.

Secure Cisco Unity

Like Cisco CallManager, Cisco Unity runs on a Windows 2000 Server operating system. Protecting Windows for Unity, including the application of security patches and hot-fixes and the installation of Cisco Security Agent and anti-virus software, is similar to the security measures described in this chapter for Cisco CallManager. For specific security guidelines and recommendations for Cisco Unity, refer to the Cisco Unity Design Guide, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps2237/products_implementation_design_guide_book09186a00801187ba.html

Voice Management

This section presents a brief overview of voice management and introduces specific tools used to manage the Cisco AVVID IP Telephony network.

Deployment Considerations

When deploying CiscoWorks applications, use the standard recommended device configuration options for management described in the documentation available at

http://www.cisco.com/en/US/netsol/ns110/ns106/ns107/ns167/networking_solutions_relevant_products.html

Cisco CallManager Settings

In addition to the standard device recommendations, configure the following settings in order to manage the Cisco Media Convergence Servers (running Cisco CallManager or other applications):

- SNMP RO community string
- SNMP RW community string

Also, for monitoring calls via the Network Analysis Module, configure the following on all Cisco CallManagers:

- Call Diagnostic Records enabled (provides packet loss and jitter information)
- IP Telephone Line Setting Display (internal caller ID)

Considerations for Voice Management

In CiscoWorks LAN Management Solution, the Discovery Schedule is set for every 4 hours. Change this default setting so that a major discovery does not happen during potential peak CPU time on Cisco CallManager. Decreasing this major discovery time will increase the network management traffic on the network and increase the load on the management stations. If the network is not changing topology very often, it is better to lengthen the Discovery Schedule to have less impact on Cisco CallManager.

CiscoWorks IP Telephony Environment Monitor Release 1.3 does not support Cisco CallManager Release 3.3.

Recommended Hardware and Software Combinations

Table A-1 lists commonly used hardware platforms for Cisco IP Telephony enterprise deployments, along with their recommended software versions.

Note The platforms and software versions listed in Table A-1 are not the only supportable deployment options. They represent the combinations of hardware and software that are subjected to the most extensive system-level testing. This ongoing testing is conducted using a variety of deployment models, several end-station size categories, and realistic call flows, traffic patterns, and use cases. For information on other possible hardware and software options for IP Telephony, see Gateway Selection, page 3-1.

Table A-1 Recommended Hardware and Software

IP Telephony Component	Software Release
Cisco CallManager	3.3(2) spB
Cisco IP Phones: 7910, 7935, 7940, and 7960	Bundled with Cisco CallManager software
Cisco Emergency Responder (CER)	1.1(4)
Cisco Customer Response Applications (CRA): IP Integrated Contact Distribution (ICD) IP Interactive Voice Response (IVR)	2.2(5) spA
Cisco Unity	4.0(1)
Cisco Unity TAPI Service Provider (TSP)	7.0(1)
Vivinet Manager	2.1 sp1
Cisco Integrated Communications System (ICS) 7750 Multiservice Route Processor (MRP)	12.2(8)YN
Cisco ICS 7750 Controller	2.6.0

Table A-1 Recommended Hardware and Software (continued)

IP Telephony Component	Software Release
Cisco Gateways: 3745 3725 3660 2691 2620XM 1760 VG200	12.2(13)T1
Cisco VG248 Gateway	1.2(1)
Cisco Catalyst 6608 and 6624 Gateways	Bundled with Cisco CallManager software
Cisco Catalyst 6506 Core Switch	6.4.1
Cisco Catalyst 6506 Multilayer Switch Feature Card (MSFC)	12.1(11b)E11
Cisco Catalyst 6509 Access Switch	6.4.1
Cisco Catalyst 4006 Access Switch	7.5.1
Cisco Catalyst 3524 Access Switch	12.0(5)WC5
Cisco Core Routers: 7206 3725 2620 1760	12.2(13)T1

For the most recent information on recommended hardware platforms and software releases, frequently refer to the online documentation at the following location:

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/ccmcomp.htm

Symbols

Numerics

A

B

C

D

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z